

RESULT 3

T49173
hypothetical protein T20N10.250 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000

C:Accession: T49173
R:DiAngelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000

A:Reference number: 225017
A:Accession: T49173

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-517 <DNA>
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250

A:Experimental source: cultivar Columbia; BAC clone T20N10

C:Genetics:
A:Gene: ATSP:T20N10.250

A:Map position: 3
A:Introns: 312/3; 359/3; 444/3

C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match 91.7%; Score 100; DB 2; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 444 KKKKKKKKKKKKKKKKKKK 463

RESULT 4

T18513
hypothetical protein C0845c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T18513
R:Lawson, D.; Bowman, S.; Barrett, B.
submitted to the EMBL Data Library, August 1997

A:Reference number: 218935
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-166 <LAW>

A:Cross-references: EMBL:Z98551; PIDN:CAB1123.2

C:Genetics:
A:Map position: 3

A:Introns: 19/1
A>Note: C0845c

Query Match 85.3%; Score 93; DB 2; Length 166;
Best Local Similarity 90.0%; Pred. No. 0.0057;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 37 KKKKKKKKKKKKKKKKKKK 56

RESULT 5

F1619
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: F1619
R:Gardner, M.J.; Tettegin, H.; Carnuci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Peterson, M.; Salberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A17600; MUID:99021743; PMID:9804551

A:Accession: F1619
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-483 <GAR>
A:Cross-references: GB:AE001382; GB:AE001362; NID:G3845130; PIDN:AACT1836.1; PID:G3845131
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0235w

Query Match 82.6%; Score 90; DB 2; Length 483;
Best Local Similarity 85.0%; Pred. No. 0.026;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 449 KKKKKKKKKKKKKKKKKKK 468

RESULT 6

C86477
protein F1504.29 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86477
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Con, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C86477
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-107 <STD>

A:Cross-references: GB:AE005172; NID:G8778346; PIDN:AAF79354.1; GSPDB:GN00141

C:Genetics:
A:Gene: F1504.29

A:Map position: 1

Query Match 79.8%; Score 87; DB 2; Length 107;
Best Local Similarity 94.4%; Pred. No. 0.019;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 19
DB 29 KKKKKKKKKKKKKKKKKKK 46

RESULT 7

A48455
acidic phosphoprotein PC5MA1q - Plasmodium chabaudi

C:Species: Plasmodium chabaudi
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998

C:Accession: A48455
R:Deleersnyder, W.; Prasomitti, P.; Tungpradubkul, S.; Hendrix, D.; Hamers-Casterman, C.

Mol. Biochem. Parasitol. 56, 59-68, 1992

A:Title: Structure of a Plasmodium chabaudi acidic phosphoprotein that is associated with

A:Reference number: A48455; MUID:93116806; PMID:1475002

A:Accession: A48455
A:Status: preliminary

A:Molecule type: nucleic acid
A:Residues: 1-441
A:Cross-references: GB:M95789; NID:G160602; PID:G160603

A:Experimental source: IP-PC1/C
A>Note: sequence extracted from NCBI backbone (NCBIN:121415, NCBI:P:121416)

Keywords: phosphoprotein
Query Match 77.1%; Score 84; DB 2; Length 441;
Best Local Similarity 80.0%; Pred. No. 0.089;
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:
Db 394 KKKKKKKKKKKKKKKKK 413

RESULT 8

hypochemical protein C0425w - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T18440
R/Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z18935
A/Accession: T18440
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-4550 <LAW>
A/Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CA11121.1
C/Genetics:
A/Map position: 3
A/Note: C0425w

Query Match 74.3%; Score 81; DB 2; Length 4550;
Best Local Similarity 44.2%; Pred. No. 0.77;
Matches 19; Conservative 2; Mismatches 0; Indels 22; Gaps 1;

OY 1 CCKK-----KKKKKKKKKKKKKKKK 21
|||||:|||||:
Db 132 CCKKNVFNINIKNYFENEKYOTNNIKKKKKKKKKKKKKKK 174

RESULT 9

hypochemical protein C0560c - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T18452
R/Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A/Reference number: Z18937
A/Accession: T18452
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-784 <LAW>
A/Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332545; PIDN:CA15594.1
C/Genetics:
A/Map position: 3
A/Note: C0560c

Query Match 73.4%; Score 80; DB 2; Length 784;
Best Local Similarity 75.0%; Pred. No. 0.31;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:
Db 360 EKKKKKKKKKKKKKKKK 379

RESULT 10

hypochemical protein DKFZp761B2423.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C/Accession: T50609
R/Boecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A/Reference number: Z25143
A/Accession: T50609
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-529 <AAA>
A/Cross-references: EMBL:AL359564

A/Experimental source: adult amygdala; clone DKFZp761B2423
C/Genetics:
A/Note: DKFZp761B2423.1

Query Match 71.6%; Score 78; DB 2; Length 529;
Best Local Similarity 80.0%; Pred. No. 0.37;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:
Db 464 KKKKKKKKKKKKKKKKK 483

RESULT 11

SAR DNA-binding protein-1 - garden pea
C/Species: Pisum sativum (garden pea)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 02-Jun-2000
C/Accession: T06377
R/Hatton, D.; Gray, J.C.
submitted to the EMBL Data Library, April 1998
A/Description: cDNA encoding a pea SAR DNA-binding protein that shows homology to nucleol
A/Reference number: Z15637
A/Accession: T06377
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-560 <HAT>
A/Cross-references: EMBL:AF061962; NID:g3132695; PIDN:AAC16330.1; PID:g3132696
C/Genetics:
A/Gene: SARBP-1
C/Superfamily: garden pea SAR DNA-binding protein

Query Match 71.6%; Score 78; DB 2; Length 560;
Best Local Similarity 75.0%; Pred. No. 0.38;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:
Db 463 KKKKKKKKKKKKKKKKK 482

RESULT 12

T42727
proliferation potential-related protein - mouse
C/Species: Mus musculus (house mouse)
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
C/Accession: T42727
R/Witte, M.M.; Scott, R.B.
submitted to the EMBL Data Library, November 1998
A/Reference number: Z22246
A/Accession: T42727
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1560 <WIT>
A/Cross-references: EMBL:U83913; NID:g3858884; PID:g3858885; PIDN:AAC72432.1
A/Experimental source: strain Balb/c
C/Genetics:
A/Gene: P2P-R
C/Function:
A/Description: involved in hnRNP association and Rb1 binding
C/Superfamily: RING finger homology
P:57-107/Domain: RING finger homology <RRN>

Query Match 71.6%; Score 78; DB 2; Length 1560;
Best Local Similarity 80.0%; Pred. No. 0.74;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:
Db 1497 KKKKKKKKKKKKKKKKK 1516

RESULT 13

JG7219
 nuclear protein SR-25 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
 C:Accession: JG7219
 R:Saahara, K., Yemaoka, T., Moritani, M., Tanaka, M., Iwahana, H., Yoshimoto, K., Miyag
 Biochem. Biophys. Res. Commun. 269, 444-450, 2000
 A:Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25
 A:Reference number: JG7219, MUID:20175222, PMID:10708573
 A:Accession: JG7219
 A:Molecule type: mRNA
 A:Reidsides: 1-229 <SAS>
 A:Cross-references: DDBJ:AB035363; NID:G7619895; PIDN:BA94743.1; PID:G7619896
 A:Experimental source: M1N6 cell line
 C:Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine r
 A:splicing factors.
 C:Keywords: nucleus; RNA processing

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 4.73239 Seconds

(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-1
Perfect score: 109
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	77.1	441	1	PHPA_PLACH
2	76	69.7	511	1	NOPS_YEAST
3	75	68.8	474	1	CBP5_SCHPO
4	75	68.8	504	1	SIK1_YEAST
5	75	68.8	2404	1	SON_MOUSE
6	75	68.8	2426	1	SON_HUMAN
7	74	67.9	414	1	Y694_METUA
8	74	67.9	474	1	CBP5_KLUJA
9	74	67.9	483	1	CBP5_YEAST
10	74	67.9	726	1	BRD3_HUMAN
11	72	66.1	479	1	CBP5_CANAL
12	72	66.1	678	1	GARP_PLAIF
13	71	65.1	351	1	CG79_HUMAN
14	71	65.1	686	1	CNG1_HUMAN
15	70.5	64.7	1411	1	TCOF_HUMAN
16	70	64.2	233	1	YD08_YEAST
17	70	64.2	599	1	HM21_HUMAN
18	70	64.2	684	1	CNG1_RAT
19	70	64.2	683	1	CNG1_MOUSE
20	69.5	63.8	534	1	NOPS_RAT
21	69.5	63.8	2231	1	SEN1_YEAST
22	69	63.3	142	1	YMH8_YEAST
23	69	63.3	167	1	YK20_YEAST
24	69	63.3	723	1	SSRP_DROME
25	69	63.3	843	1	BLVR_BOVIN
26	68.5	62.8	724	1	Y061_CAEEL
27	68	62.4	523	1	DBP3_YEAST
28	68	62.4	1178	1	MANA_YEAST
29	67	61.5	118	1	YQ93_CAEEL
30	67	61.5	690	1	CNG1_BOVIN
31	67	61.5	691	1	CNG1_CANFA
32	67	61.5	1002	1	IF2P_YEAST
33	67	61.5	1220	1	IF2P_HUMAN

34	67	61.5	1362	1	BRD4_HUMAN	O60885 homo sapien
35	67	61.5	2280	1	YCF2_OENHO	O9mf22 oenothera h
36	66	60.6	481	1	CBP5_BEMNI	O43100 emeticella
37	66	60.6	487	1	CBP5_ASFPU	O43102 aspergillus
38	66	60.6	667	1	YK11_SCHPO	O13796 schizosacch
39	66	60.6	1153	1	AAD1_HUMAN	O14617 homo sapien
40	66	60.6	1240	1	YJ41_YEAST	P53935 saccharomyc
41	65.5	60.1	508	1	NO60_DROME	O44081 drosophila
42	65.5	60.1	514	1	DKC1_HUMAN	O60832 homo sapien
43	65	59.6	217	1	KSL1_HYDAT	P38978 hydra atten
44	65	59.6	271	1	YGSW_YEAST	P53335 saccharomyc
45	65	59.6	320	1	YD33_YEAST	Q12117 saccharomyc

ALIGNMENTS

RESULT 1	ID	PHPA_PLACH	STANDARD	PRT	441 AA.
AC	002752	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	01-JUN-1994 (Rel. 29, Last annotation update)				
DE	Acidic phosphoprotein precursor (50 kDa antigen).				
CN	PCEMAL				
OS	Plasmodium chabaudi.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5825;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-IP-PCI;				
RX	MEDLINE=93116806; PubMed=1475002;				
RA	Deleersnyder W., Prasomsilti P., Tungradabkul S., Hendrix D.,				
RA	Hamers-Casterman C., Hamers R.;				
RT	"Structure of a plasmodium chabaudi acidic phosphoprotein that is				
RT	associated with the host erythrocyte membrane."				
RL	Mol. Biochem. Parasitol. 56:59-68(1992)				
CC	-1- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES				
CC	THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE				
CC	PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.				
CC	-1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE				
CC	CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.				
CC	-1- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE				
CC	THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; M95789; AAA29732.1; -				
DR	PIR; A48455; A48455.				
KW	Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.				
FT	CHAIN	1	15		OR 24 (POTENTIAL).
FT	SIGNAL	16	441		ACIDIC PHOSPHOPROTEIN.
FT	DOMAIN	186	313		16 X 8 AA TANDEM REPEATS.
FT	REPEAT	186	193		1-1.
FT	REPEAT	194	201		1-2.
FT	REPEAT	202	209		1-3.
FT	REPEAT	210	217		1-4.
FT	REPEAT	218	225		1-5.
FT	REPEAT	226	233		1-6.
FT	REPEAT	234	241		1-7.
FT	REPEAT	242	249		1-8.
FT	REPEAT	250	257		1-9.
FT	REPEAT	258	265		1-10.
FT	REPEAT	266	273		1-11.
FT	REPEAT	274	281		1-12.
FT	REPEAT	282	289		1-13.

FT REPEAT 290 297 1-14.
 FT REPEAT 298 305 1-15.
 FT REPEAT 306 313 1-16.
 FT DOMAIN 353 370 2 X 9 AA TANDEM REPEATS.
 FT REPEAT 353 360 2-1.
 FT REPEAT 361 368 2-2.
 FT DOMAIN 371 417 LYS-RICH (BASIC).
 FT CAROAND 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROAND 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 441 AA, 49708 MW, DB5E83E795EE7E5 CR664;

Query Match 77.1%; Score 84; DB 1; Length 441;
 Best Local Similarity 80.0%; Pred. No. 0.032;
 Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 394 KKKKKKKKKKKKKKKKK 413

RESULT 2
 NOPS_YEAST STANDARD; PRT; 511 AA.
 AC Q12499;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleolar protein NOPS8 (Nucleolar protein NOPS).
 GN NOPS8 OR NOPS OR YOR310C OR O6108.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=303;
 RA Pearson B.M., Herrando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 RT pre-18S rRNA processing in yeast."
 RT "Bio. Chem. 273:16453-16463(1998).
 CC -1- FUNCTION: REQUIRED FOR PRE-18S RNA PROCESSING. MAY BIND
 CC MICROTUBULES.
 CC -1- SUBUNIT: INTERACTS WITH NOPS6 AND NOP1.
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
 CC -1- SIMILARITY: BELONGS TO THE NOPS/NOPS6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X90565; CA62165.1; -.
 CC EMBL, 275217; CA99630.1; -.
 CC EMBL, AF056070; AAC39484.1; -.
 CC PIR, S58322; S58322.
 CC SGI, S0005837; NOPS8.
 CC GO, GO:0005735; C:sma11 nucleolar ribonucleoprotein complex; IPI.
 CC GO, GO:0003754; F:chaperone activity; NAS.
 CC GO, GO:001069; F:snRNA binding activity; IIA.
 CC GO, GO:0030490; P:processing of 20S pre-rRNA; IPI.
 CC GO, GO:0006608; P:snNP protein-nucleus import; NAS.
 CC InterPro, IPR002687; NOP.
 CC Pfam, PF01798; NOP; 1.
 CC Prodom, PD004104; NOP; 1.

KM Ribosome biogenesis; Nuclear protein; rRNA processing.
 FT DOMAIN 441 511 ASP/GLU/LYS-RICH.
 SQ SEQUENCE 511 AA, 56956 MW, 8A2889448B2A19E2 CR664;

Query Match 69.7%; Score 76; DB 1; Length 511;
 Best Local Similarity 70.0%; Pred. No. 0.21;
 Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 480 KKKKKKKKKKKKKKKKK 499

RESULT 3
 CBP5_SCHPO STANDARD; PRT; 474 AA.
 AC Q14007;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Centromere/microtubule binding protein cbf5 (centromere-binding factor
 5) (Nucleolar protein cbf5).
 GN CBP5 OR SPAC29A4.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA MEDLINE=21846401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felkell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris P., Hildalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Uegels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolke J., Aert R., Robben J., Glynn B.,
 RA Wellens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Ussery D., Bartell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 413:871-880(2002).
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CENTROMERE DNA-CBP3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC -1- SIMILARITY: Contains 1 PUA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC EMBL; Z97210; CAB10131.1; -.
DR PIR; T38485; T38485.
DR GenBank; SP000000; SP000000.04c; -.
DR InterPro; IPR004602; CbU5.
DR InterPro; IPR004278; PUA.
DR InterPro; IPR004501; TUB_N.
DR InterPro; IPR004521; Unchar_dom_2.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; TUB_N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGR; TIGR00425; CBF5; 1.
DR TIGR; TIGR00451; unchar_dom_2; 1.
DR PROSITE; PS00890; PUA; 1.
KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 271 346 PUA.
FT 434 468 7 x 3 AA APPROXIMATE TANDEM REPEATS OF
FT REPEAT 443 445 1.
FT REPEAT 450 452 2.
FT REPEAT 454 456 3.
FT REPEAT 457 459 4.
FT REPEAT 460 462 5.
FT REPEAT 463 465 6.
FT REPEAT 466 468 7.
SQ SEQUENCE 474 AA; 53110 MW; B8C9896C5FAEB71 CRC64;

Query Match 68.8%; Score 75; DB 1; Length 474;
Best Local Similarity 73.7%; Pred. No. 0.24;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 20
Db 454 KKKKKKKKKKKKKKKKKKK 472

RESULT 4
ID SIK1_YEAST STANDARD; PRT; 504 AA.
AC Q12460;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE SIK1 protein (Nucleolar protein NOP56).
GN SIK1 OR NOP56 OR YLR197W OR I6167.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=5288C / YPH1;
RX MEDLINE=96040178; PubMed=7547500;
RA Johnson M., Hillier L., Riles L., Albermann K., Andre B., Anseorge W.,
RA Benes V., Brueckner M., Delius H., Dubois A., Duesterhoeft A.,
RA Enlian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Nelzel D., Hilbert H., Hilger F., Kleine K., Koelter P.,
RA Louis E.-J., Messenguy F., Mews H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Newtlich U., Oberwieser B., Piravandi E., Pohl T.M.,
RA Poterelle D., Purnell B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scheu B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wamboldt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hant U., Hohnselt J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
```

```

RL Nature 387:87-90 (1997).
RN [3]
RP CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=98038777; PubMed=9372940;
RA Gauthier T., Berges T., Tollervy D., Hurt E.;
RT "Nucleolar KEX/D repeat proteins Nop56p and Nop58p interact with Nop1p
and are required for ribosome biogenesis."
RL Mol. Cell. Biol. 17:7088-7098 (1997).
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.
CC -1- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```

```

DR EMBL; U20237; AAC49066.1; -.
DR EMBL; U14913; AAB67431.1; -.
DR PIR; S48550; S48550.
DR SGD; S0004187; SIK1.
DR GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
DR InterPro; IPR002687; NOP.
DR Pfam; PF01798; NOP; 1.
DR ProDom; PD004104; NOP; 1.
KW Ribosome biogenesis; Nuclear protein.
FT DOMAIN 443 504 ASP/GLU/LYS-RICH.
FT MUTAGEN 333 333 V->A: REDUCED GROWTH RATE AT ALL
TEMPERATURES; WHEN ASSOCIATED WITH R-385.
FT MUTAGEN 355 355 Y->C: AT 37 DEGREES, GROWTH SLOWS AFTER 6
TO 8 HOURS AND CELL DIVISION STOPS AFTER
20 HOURS.
FT MUTAGEN 385 385 M->R: REDUCED GROWTH RATE AT ALL
TEMPERATURES; WHEN ASSOCIATED WITH A-333.
SQ SEQUENCE 504 AA; 56864 MW; F8522A5870EF4842 CRC64;

Query Match 68.8%; Score 75; DB 1; Length 504;
Best Local Similarity 70.0%; Pred. No. 0.25;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 465 KKKKKKKKKKKKKKKKKKK 464

RESULT 5
ID SON_MOUSE STANDARD; PRT; 2404 AA.
AC Q9QX47; Q9QX47; Q9QX47; Q9QX47; Q9QX47;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last sequence update)
DE SON protein.
GN SON.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/SV;
RX MEDLINE=20408886; PubMed=10950926;
RA Wynn S.L., Fisher R.A., Pegg C., Price M., Liu Q.Y., Khan I.M.,
RA Zammit P., Dadgar K., Mazzanti W., Kessling A., Lee J.S., Bulwella L.;
RT "Organization and conservation of the GART/SON/DONSON locus in mouse
and human genomes."
RL Genomics 68:57-62 (2000).
RN [2]
```

RT SEQUENCE OF 1-116 FROM N.A.
 RC STAIN-NC57BL/6J: TISSUE=hippocampus, Small intestine, and Tongue;
 EX MEDLINE=21085560; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,
 RA Blake Y., Boffelli D., Bojunga M., Carnici P., de Bonaldo M.F.,
 RA Brownstein M.J., Bulc C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima T., Mazzarelli U., Mombertis P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Witting L.,
 RA Wyszewski-Boris A., Yoshida K., Haegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 CC -1- FUNCTION: Transcriptional repressor. Binds to the consensus DNA
 CC sequence: 5'-GATG[AT]AN[CG] (AG)CC-3'. Might protect cells from
 CC apoptosis. Might be involved in pre-mRNA splicing (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=G9QX47-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=G9QX47-2; Sequence=VSP_004416, VSP_004417;
 CC -1- TISSUE SPECIFICITY: Widely expressed.
 CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3
 CC regions.
 CC -1- SIMILARITY: Contains 1 G-patch domain.
 CC -1- SIMILARITY: Contains 1 DBM (double-stranded RNA-binding) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: AF193606; AAF23120.1; -
 DR EMBL: AF193595; AAF23120.1; JOINED.
 DR EMBL: AF193596; AAF23120.1; JOINED.
 DR EMBL: AF193597; AAF23120.1; JOINED.
 DR EMBL: AF193598; AAF23120.1; JOINED.
 DR EMBL: AF193599; AAF23120.1; JOINED.
 DR EMBL: AF193600; AAF23120.1; JOINED.
 DR EMBL: AF193601; AAF23120.1; JOINED.
 DR EMBL: AF193602; AAF23120.1; JOINED.
 DR EMBL: AF193603; AAF23120.1; JOINED.
 DR EMBL: AF193604; AAF23120.1; JOINED.
 DR EMBL: AF193605; AAF23120.1; JOINED.
 DR EMBL: AF193607; AAF23121.1; -
 DR EMBL: AK019312; BAB31659.1; -
 DR EMBL: AK019081; BAB31536.1; -
 DR EMBL: AK008478; BAB25691.1; -
 DR EMBL: AK008256; BAB25562.1; -
 DR MGD: MGI:98159; Son.
 DR GO: GO:0005515; P:protein binding activity; IPT.
 DR InterPro: IPR001159; DS_RBD.
 DR InterPro: IPR000467; G_patch.
 DR Pfam: PF00035; dnm; 1.
 DR Pfam: PF01585; G_patch; 1.
 DR SMART: SM00443; G_patch; 1.
 DR PROSITE: PS0137; DS_RBD; 1.
 DR PROSITE: PS0174; G_patch; 1.

Query Match	Best Local Similarity	69.8%;	Score 75;	DB 1;	Length 2404;
Matches	15;	Conservative	2;	Mismatches	3;
				Indels	0;
				Gaps	0;
Qy	2	XXXXXXXXXXXXXXXXXXXX	21		
Db	109	KXHKXKXKXKXKXKXKXKX	128		
RESULT 6					
SON_HUMAN	STANDARD;	PRT;	2426	AA.	
ID	SON_HUMAN				
AC	P15583; Q14487; Q95981; Q14120; Q9H7B1; Q9P070; Q9P072; Q9UKP9;				
AC	Q9UPY0;				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	SON protein (SON3) (Negative regulatory element-binding protein) (NRE-binding protein) (DBP-5) (Bax antagonist selected in saccharomyces 1) (BASS1) (Protein C2orf50).				
DE	SON OR NREBP OR DBP5 OR C21ORF50 OR KIA1019.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS A; B; C; D; E AND F).				
RX	MEBLINE=21564202; PubMed=11707072;				

RA Raymond A., Friedli M., Neergaard Henriksen C., Chapot F.,
 RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
 RA Antonarakis S.E.;
 RT "From PREDS and open reading frames to cDNA isolation: revisiting the
 RT human chromosome 21 transcription map.";
 RL Genomics 78:46-54(2001).
 [12]
 RN SEQUENCE FROM N.A. (ISOFORM G).
 RP TISSUE=Liver;
 RC MEDLINE=21316479; PubMed=11306577;
 RX Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,
 RA Ting L.-P.;
 RT "Transcription repression of human hepatitis B virus genes by negative
 RT regulatory element-binding protein/SOX.";
 RL J. Biol. Chem. 276:24059-24067(2001).
 [13]
 RN SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
 RP TISSUE=Placenta;
 RC Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,
 RA Vitale L., Giannone S., Carinci P., Zannotti M.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 [14]
 RN SEQUENCE OF 1-130 FROM N.A.
 RP TISSUE=Smooth muscle;
 RC Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
 RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Oca T., Suzuki Y.,
 RA Ohsayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isogai T., Sugano S.;
 RT "NEBO human cDNA sequencing project.";
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 [15]
 RN SEQUENCE OF 1-114 FROM N.A.
 RP TISSUE=Blood;
 RC Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
 RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
 RT "Human partial CDS from cd34+ stem cells.";
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 [16]
 RN SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
 RP TISSUE=Brain;
 RC MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
 RA Tanaka A., Kozani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 [17]
 RN SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
 RP MEDLINE=92049296; PubMed=1944255;
 RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
 RA Prasolov V.S.;
 RT "Identification of a protein product of a novel human gene SON and
 RT the biological effect upon administering a changed form of this gene
 RT into mammalian cells.";
 RL Mol. Biol. (Mosk) 25:731-740(1991).
 [18]
 RN SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
 RP TISSUE=Placenta;
 RC MEDLINE=93062885; PubMed=1435774;
 RX Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
 RA "The human son gene: the large and small transcripts contains various
 RT 5'-terminal sequences.";
 RL Mol. Biol. (Mosk) 26:807-812(1992).
 [19]
 RN SEQUENCE OF 1009-1131 FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=93062884; PubMed=1435773;
 RX Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,
 RA Chumakov I.M.;
 RT "Coding part of the son gene small transcript contains four areas of
 RT complete tandem repeats."

RL Mol. Biol. (Mosk) 26:793-806(1992).
 [10]
 RN SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).
 RP MEDLINE=93048367; PubMed=1424386;
 RX Mattioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,
 RA Lee J.S.;
 RT "A cDNA clone for a novel nuclear protein with DNA binding
 RT activity.";
 RL Chromosoma 101:618-624(1992).
 [11]
 RN SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).
 RP MEDLINE=89039788; PubMed=3054499;
 RX Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;
 RT "Decoding of the primary structure of the son3 region in human
 RT genome: identification of a new protein with unusual structure and
 RT homology with DNA-binding proteins.";
 RL Mol. Biol. (Mosk) 22:794-801(1988).
 [12]
 RN SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).
 RP TISSUE=Cerebellum;
 RC MEDLINE=99439804; PubMed=10509013;
 RX Greenhalf W., Lee J., Chaudhuri B.;
 RT "A selection system for human apoptosis inhibitors using yeast.";
 RL Yeast 15:1307-1321(1999).
 CC -1- FUNCTION: Represses hepatitis B virus (HBV) core promoter activity
 CC and transcription of HBV genes and production of HBV virions.
 CC Binds to the consensus DNA sequence: 5'-GAGTGTAN[CG]AG[CC-3'.
 CC Might protect cells from apoptosis. Might be involved in pre-mRNA
 CC splicing (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=10;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=F;
 CC IsoId=P18583-1; Sequence=Displayed;
 CC Name=A;
 CC IsoId=P18583-2; Sequence=VSP_004401, VSP_004402, VSP_004403;
 CC Name=B;
 CC IsoId=P18583-3; Sequence=VSP_004404, VSP_004405;
 CC Name=C;
 CC IsoId=P18583-4; Sequence=VSP_004406, VSP_004407;
 CC Name=D;
 CC IsoId=P18583-5; Sequence=VSP_004403;
 CC Name=E;
 CC IsoId=P18583-6; Sequence=VSP_004408, VSP_004409;
 CC Name=G;
 CC IsoId=P18583-7; Sequence=VSP_004410;
 CC Name=H;
 CC IsoId=P18583-8; Sequence=VSP_004411, VSP_004412;
 CC Name=I;
 CC IsoId=P18583-9; Sequence=VSP_004413;
 CC Name=J;
 CC IsoId=P18583-10; Sequence=VSP_004414, VSP_004415;
 CC -1- TISSUE SPECIFICITY: Widely expressed, with the higher expression
 CC seen in leukocyte and heart.
 CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3
 CC regions.
 CC -1- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor
 CC SFRS2/SC-35.
 CC -1- SIMILARITY: Contains 1 G-patch domain.
 CC -1- SIMILARITY: Contains 1 DBM (double-stranded RNA-binding) domain.
 CC -1- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN DUE
 CC TO A FRAMESHIFT.
 CC -1- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN
 CC DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>)

or send an email to license@ib-sib.ch).

CC EMBL; AF380179; AAL34497.1; -
 DR EMBL; X63753; CAA35282.1; ALT_FRAME.
 DR EMBL; M36438; AAA36624.1; -
 DR EMBL; AF380180; AAL34498.1; -
 DR EMBL; AF380181; AAL34499.1; -
 DR EMBL; AF380182; AAL34500.1; -
 DR EMBL; AF380183; AAL34501.1; -
 DR EMBL; AF380184; AAL34502.1; -
 DR EMBL; AY025895; AAK07692.1; -
 DR EMBL; AF435977; AAL30810.1; -
 DR EMBL; X63751; CAC69885.1; -
 DR EMBL; AB028942; BAA82971.1; -
 DR EMBL; X63071; CAA44793.1; ALT_FRAME.
 DR EMBL; AF139897; AAD50078.1; -
 DR EMBL; AK024752; BAB14985.1; -
 DR EMBL; AF161428; AAF28988.1; -
 DR EMBL; AF161430; AAF28990.1; -
 DR Genew; HGNC:11183; SON.
 DR GK; P18583; -
 DR MIM; 182465; -
 DR GO; GO:0008189; F.apoptosis inhibitor activity; IDA.
 DR GO; GO:0003677; F.DNA binding activity; TAS.
 DR GO; GO:0006916; P.apoptosis; IDA.
 DR InterPro; IPR001159; G_RBD.
 DR InterPro; IPR000467; G_patch.
 DR Pfam; PF00035; dsm; 1.
 DR Pfam; PF01585; G_patch; 1.
 DR SMART; SM00358; DSRM; 1.
 DR SMART; SM00443; G_patch; 1.
 DR PROSITE; PS0137; DS_RBD; 1.
 DR PROSITE; PS0174; G_PATCH; 1.
 DR RNA-binding; DNA-binding; Nuclear protein; Repeat;
 KW Alternative splicing.
 FT DOMAIN 726 895
 [NSG]-(TS)-MDSQM.
 17 X 10 AA TANDEM REPEATS OF L-A-[ST]-

Query Match 68.8%; Score 75; DB 1; Length 2426;
 Best Local Similarity 75.0%; Pred. No. 0.86;
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
 Db 109 KKKKKKKKKKKKKKKKKKKKK 128

RESULT 7
 Y694 METUA STANDARD; PRT; 414 AA.
 AC 058105;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein Mj0694.
 GN Mj0694.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxId=2190;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weissflog K.G., Merrick J.M., Glöck A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Cutlerback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.",
 RL Science 273:1058-1073 (1996).
 CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ibs-sib.ch/announce/>
 CC or send an email to license@ib-sib.ch).

CC EMBL; U67516; AAB98689.1; -
 DR PIR; F64386; F64386.
 DR TIGR; Mj0694; -
 DR InterPro; IPR002687; NOP.
 DR Pfam; PF01798; NOP; 1.
 DR Prodom; PD004104; NOP; 1.
 DR Hypothetical protein; Complete proteome.
 KW Domain 349 414 ASP/GLU/LYS-RICH.
 FT SEQUENCE 414 AA; 47799 MW; A9092EFC3C82C407 CRC64;

Query Match 67.9%; Score 74; DB 1; Length 414;
 Best Local Similarity 75.0%; Pred. No. 0.27;
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
 Db 377 KKKKKKKKKKKKKKKKKKKKK 396

RESULT 8
 CBP5_KLUUA STANDARD; PRT; 474 AA.
 ID CBP5_KLUUA
 AC 013473;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Centromere/microtubule binding protein CBP5 (Centromere-binding factor
 DE 5) (Nucleolar protein CBP5).
 GN CBP5.
 OS Kluveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.
 OX NCBI_TaxId=26985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JBD100;
 RX MEDLINE=98144788; PubMed=9483794;
 RA Winkler A.A., Bobok A., Zonneveld B.J.M., Steensma H.Y.,
 RA Hooykaas P.J.J.;
 RA "The lysine-rich C-terminal repeats of the centromere-binding factor
 RT 5 (Cbif) of Kluveromyces lactis are not essential for function.",
 RL Yeast 14:37-48 (1998).
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CENTROMERE DNA-CBP5-BINDING FACTOR AND IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
 CC SOME WAY ASSOCIATED WITH THE CBP5 110 kDa SUBUNIT (CBP3A) (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC -1- SIMILARITY: Contains 1 PUA domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ibs-sib.ch/announce/>
 CC or send an email to license@ib-sib.ch).

CC EMBL; AF008563; AAC64862.1; -
 DR InterPro; IPR004802; CBP5.


```

DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; Trub_N.
DR InterPro; IPR004521; Unchar_dom_2.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; Trub_N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGRFAMs; TIGR00425; CBF5; 1.
DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR PROSITE; PS00890; PUA; 1.
KM Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 285 340 PUA.
FT DOMAIN 431 460 1.
FT REPEAT 431 463 9 x 3 AA TANDEM REPEATS OF K-K [DE].
FT REPEAT 434 436 2.
FT REPEAT 437 439 3.
FT REPEAT 440 442 4.
FT REPEAT 443 445 5.
FT REPEAT 446 448 6.
FT REPEAT 449 451 7.
FT REPEAT 452 454 8.
FT REPEAT 455 457 9.
SQ SEQUENCE 474 AA; 53630 MW; 95306EC7FEA756C CRC64;

Query Match 67.9%; Score 74; DB 1; Length 474;
Best Local Similarity 70.0%; Pred. No. 0.3;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 434 KKKKKKKKKKKKKKKKKKK 453

RESULT 9
CBF5_YEAST STANDARD; PRT; 483 AA.
ID CBF5_YEAST STANDARD; PRT; 483 AA.
AC P33322;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor 5) (Nucleolar protein CBF5) (P64').
GN CBF5 OR YLR175W OR U9470.11.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93330283; PubMed=8336724;
RA Jiang W., Middleton K., Yoon H.-J., Fouquet C., Carbon J.;
RA "An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules."
RT Mol. Cell. Biol. 13:484-4893(1993).
RL [2]
RM SEQUENCE FROM N.A.
RP STRAIN=5288C / AB972;
RC MEDLINE=97313267; PubMed=9169871;
RX Johnson M., Hillier L., Riles L., Albermann K., Andre B., Ansgore W., Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A., Eutian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K., Heuss-Neitzel D., Hilbert H., Hilger F., Kline K., Koetter P., Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D., Mueller-Auer S., Netewich U., Obermaier B., Piravandi E., Pohl T.M., Portetlelle D., Purnelle B., Recheimann S., Rieger M., Rinke M., Rose M., Scharfe M., Scherens B., Scholler P., Schwaiger C., Schwartz S., Underwood A.P., Urrutazu L.A., Vandenbol M., Verhaesselt P., Viereidels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E., Wedler H., Zimmermann F.K., Zollner A., Han J., Hohelsel J.D., Wedler E., "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII." Nature 387:87-90(1997).
RL -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A CENTROMERE DNA-CBF5-BINDING FACTOR AND IS INVOLVED IN MITOTIC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN

```

```

CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDa SUBUNIT (CBF3A).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -1- SIMILARITY: Contains 1 PUA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12351; AAA34473.1; -.
DR EMBL; U17246; BAB67463.1; -.
DR PIR; S41853; S41853.
DR SGD; S0004165; CBF5.
DR GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
DR InterPro; IPR004802; Cbf5.
DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; Trub_N.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; Trub_N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR PROSITE; PS00890; PUA; 1.
DR Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
KM Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 266 341 PUA.
FT DOMAIN 434 463 10 x 3 AA TANDEM REPEATS OF K-K [DE].
FT REPEAT 434 436 1.
FT REPEAT 437 439 2.
FT REPEAT 440 442 3.
FT REPEAT 443 445 4.
FT REPEAT 446 448 5.
FT REPEAT 449 451 6.
FT REPEAT 452 454 7.
FT REPEAT 455 457 8.
FT REPEAT 458 460 9.
FT REPEAT 461 463 10.
SQ SEQUENCE 483 AA; 54704 MW; D356B39FDC32E2D CRC64;

Query Match 67.9%; Score 74; DB 1; Length 483;
Best Local Similarity 70.0%; Pred. No. 0.3;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 434 KKKKKKKKKKKKKKKKKKK 453

RESULT 10
BRD3_HUMAN STANDARD; PRT; 726 AA.
ID BRD3_HUMAN STANDARD; PRT; 726 AA.
AC Q15059; Q92645;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 3 (RING3-like protein).
GN BRD3 OR RING3L OR KIAA0043.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP TISSUE=Bone marrow;
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S., Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RA "Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by

```

RT analysis of cDNA clones from human cell line KG-1."
 RL DNA Res. 1:223-229(1994).
 RN [2]
 RP SEQUENCE OF 363-726 FROM N.A.
 RX MEDLINE=98038990; PubMed=9373153;
 RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.,
 RT "Chromosomal localization, gene structure and transcription pattern of
 the ORF gene, a homologue of the MHC-linked RING3 gene."
 RL gene 200:177-183(1997).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- SIMILARITY: Contains 2 Bromodomains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D26362; BA05393.1; -
 DR EMBL; Z81330; CAB03630.1; -
 DR HSSP; Q92831; 1B91.
 DR Genew; HGNC:1104; BRD3.
 DR MIM; 601541; -
 DR GO; GO:0005634; C:nucleus; NAS.
 DR InterPro; IPR001487; Bromodomain.
 DR Pfam; PF00439; bromodomain; 2.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 2.
 DR PROSITE; PS00633; BROMODOMAIN 1; 2.
 DR PROSITE; PS50014; BROMODOMAIN 2; 2.
 DR Bromodomain; Repeat; Nuclear Protein.
 FT DOMAIN 56 115 BROMODOMAIN 1.
 FT DOMAIN 326 398 BROMODOMAIN 2.
 FT DOMAIN 487 555 LYS-RICH.
 FT DOMAIN 676 725 SER-RICH.
 FT CONFLICT 465 466 EL -> DV (IN REF. 2).
 FT SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;
 SQ
 Query Match 67.9%; Score 74; DB 1; Length 726;
 Best Local Similarity 70.0%; Pred. No. 0.42;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 KKKKKKKKKKKKKKKKKKKKK 21
 Db 489 KKKKKKKKKKKKKKKKKKKKK 508
 RESULT 11
 CBFS CANAL STANDARD; PRT; 479 AA.
 ID CBFS CANAL
 AC 043101;
 DT 15-JUL-1998 (Rel. 36; Created)
 DT 15-JUL-1998 (Rel. 36; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Centromere/microtubule binding protein CBFS (Centromere-binding factor
 DE 5) (Nucleolar protein CBFS).
 GN CBFS.
 OS Candida albicans (Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetaceae; mitosporic Saccharomycetaceae; Candida.
 CC NCB1_TaxID=5476;
 CC [1]
 CC SEQUENCE FROM N.A.
 RA Jiang W., Cliford J., Koltin Y.,
 RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
 CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A) (BY
 CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC -1- SIMILARITY: Contains 1 PUA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U59149; AAB94297.1; -
 DR InterPro; IPR004802; Cof5.
 DR InterPro; IPR002478; PUA.
 DR InterPro; IPR002501; TRUB_N.
 DR InterPro; IPR004521; Unchar_dom_2.
 DR Pfam; PF01472; PUA; 1.
 DR Pfam; PF01509; TRUB_N; 1.
 DR SMART; SM00359; PUA; 1.
 DR TIGRFAMs; TIGR00425; CBF5; 1.
 DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
 DR PROSITE; PS50890; PUA; 1.
 DR Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
 KW Microtubules; Repeat; Nuclear protein; DNA-binding.
 FT DOMAIN 267 342 PUA.
 FT SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;
 SQ
 Query Match 66.1%; Score 72; DB 1; Length 479;
 Best Local Similarity 70.0%; Pred. No. 0.47;
 Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Oy 2 KKKKKKKKKKKKKKKKKKKKK 21
 Db 432 KKKKKKKKKKKKKKKKKKKKK 451
 RESULT 12
 GARP PLAFF STANDARD; PRT; 678 AA.
 ID GARP PLAFF
 AC P13816;
 DT 01-JAN-1990 (Rel. 13; Created)
 DT 01-JAN-1990 (Rel. 13; Last sequence update)
 DT 15-JUL-1999 (Rel. 38; Last annotation update)
 DE Glutamic acid-rich protein precursor.
 GN GARP.
 OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 CC NCB1_TaxID=5837;
 CC [1]
 CC SEQUENCE FROM N.A.
 RX MEDLINE=89040048; PubMed=2903445;
 RA Trigila T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
 RA Kemp D.J.,
 RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
 RT acid-rich protein (GARP)."
 RL Mol. Biochem. Parasitol. 31:199-202(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J03998; AAA29605.1; -
 DR PIR; A54514; A54514.
 KM Repeat; Malaria; Antigen; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
 FT DOMAIN 120 164 15 X 3 AA TANDEN REPEATS OF K-K-X.
 FT DOMAIN 372 416 9 X APPROXIMATE TANDEN REPEATS.
 FT DOMAIN 417 441 5 X APPROXIMATE TANDEN REPEATS.

FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU
SQ SEQUENCE 678 AA; 80551 MW; 2A6F856064965A9E CRC64;
Query Match 66.1%; Score 72; DB 1; Length 678;
Best Local Similarity 70.0%; Pred. No. 0.62;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 2 KKKKKKKKKKKKKKKKKKK 21
DB 132 KKKKKKKKKKKKKKKKKKK 151
RESULT 13
CG79_HUMAN STANDARD; PRT; 351 AA.
AC Q9Y388;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein CGI-79.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20272150; PubMed=10810093;
RX Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RL Caenorhabditis elegans by comparative proteomics."
CC Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF151837; AAD34074.1; -
DR HSSP; P26368; 202P.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KM Hypothetical protein; RNA-binding.
FT DOMAIN 36 114 RNA-BINDING (RRM).
SQ SEQUENCE 351 AA; 39675 MW; 7B6E882D68192EBE CRC64;
Query Match 65.1%; Score 71; DB 1; Length 351;
Best Local Similarity 73.7%; Pred. No. 0.46;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 2 KKKKKKKKKKKKKKKKKKK 20
DB 156 KKKKKKKKKKKKKKKKKKK 174
RESULT 14
CG1_HUMAN STANDARD; PRT; 686 AA.
AC P29973; Q16279; Q16485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG1)
DE (Cyclic nucleotide-gated cation channel alpha 1) (Cyclic nucleotide-gated
DE channel, photoreceptor) (Cyclic nucleotide-gated cation channel 1)

DE (Rod photoreceptor cGMP-gated channel alpha subunit).
GN CNG1 OR CNG1 OR CNGC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92210603; PubMed=1372902;
RX Pittler S.J., Lee A.K., Altherr M.R., Howard T.A., Seldin M.F.,
RT Hurwitz R.L., Masmuth J.J., Baehr W.;
RT "Primary structure and chromosomal localization of human and mouse
RT rod photoreceptor cGMP-gated cation channel."
RL J. Biol. Chem. 267:6257-6262(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=92356211; PubMed=1379636;
RA Dhaliwal R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,
RA Yau K.-W., Nathans J.;
RT "Human rod photoreceptor cGMP-gated channel: amino acid sequence,
RT gene structure, and functional expression."
RL J. Neurosci. 12:3248-3256(1992).
RN [3]
RP SEQUENCE OF 313-573 FROM N.A.
RA MEDLINE=95175019; PubMed=7532814;
RX Dietler W., Biel M., Flocke V., Hofmann F.;
RT "Expression of cyclic nucleotide-gated cation channels in non-sensory
RT tissues and cells."
RL Neuropharmacology 33:1275-1282(1994).
RN [4]
RP VARIANT ARR PHE-316, AND VARIANTS GLN-28 AND ASN-114.
RX MEDLINE=96036047; PubMed=7479749;
RA Dryja T.P., Finn J.T., Peng Y.-W., McGee T.L., Berson E.L., Yau K.-W.;
RT "Mutations in the gene encoding the alpha subunit of the rod
RT cGMP-gated channel in autosomal recessive retinitis pigmentosa."
RL Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).
CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC COUPLED CASCADE USING cGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO AN OPENING OF THE CATION
CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC PHOTORECEPTORS.
CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC
CC COMPLEX WITH CNG4.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
CC -1- DISEASE: Defects in CNGA1 are a cause of autosomal recessive
CC retinitis pigmentosa (ARRP) [MIM:123825], a disease that leads to
CC degeneration of retinal photoreceptor cells.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC FAMILY.
CC -1- DATABASE: NAME=Mutations of the CNGA1 gene;
CC NOTE=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/cngalmut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84741; AAA52010.1; ALT_INIT.
DR EMBL; S42457; AAB22778.1; -
DR EMBL; S76062; AAD14206.1; -
DR Genew; HGNC:2148; CNGA1.
DR MIM; 123825; -
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0006832; P:molecule transport; TAS.
DR GO; GO:0007601; P:vision; TAS.
DR InterPro; IPR000595; cGMP_binding.


```
DR EMBL; U79646; AAB40722.1; JOINED.
DR EMBL; U79647; AAB40722.1; JOINED.
DR EMBL; U79648; AAB40722.1; JOINED.
DR EMBL; U79649; AAB40722.1; JOINED.
DR EMBL; U79650; AAB40722.1; JOINED.
DR EMBL; U79651; AAB40722.1; JOINED.
DR EMBL; U79652; AAB40722.1; JOINED.
DR EMBL; U79653; AAB40722.1; JOINED.
DR EMBL; U79654; AAB40722.1; JOINED.
DR EMBL; U79655; AAB40722.1; JOINED.
DR EMBL; U79656; AAB40722.1; JOINED.
DR EMBL; U79657; AAB40722.1; JOINED.
DR EMBL; U79658; AAB40722.1; JOINED.
DR Genew; HGNC:11654; TCOF1.
DR MIM; 606847; -.
DR MIM; 154500; -.
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0005215; F:transporter activity; TAS.
DR GO; GO:0001501; P:skeletal development; TAS.
DR InterPro; IPR006594; LISH.
DR InterPro; IPR003993; treacle.
DR Pfam; PF03546; treacle; 3.
DR PRINTS; PR01503; TREACLE.
DR SMART; SMO0667; LISH; 1.
DR PROSITE; PS00896; LISH; 1.
KW Disease mutation; Polymorphism.
FT DOMAIN 6 38 LISH.
FT DOMAIN 89 97 POLY-GLU.
FT DOMAIN 204 207 POLY-SER.
FT DOMAIN 616 619 POLY-SER.
FT DOMAIN 919 924 POLY-SER.
FT DOMAIN 1285 1289 POLY-LYS.
FT DOMAIN 1375 1375 POLY-LYS.
FT DOMAIN 1398 1405 POLY-LYS.
FT VARIANT 53 53 W -> R (in TCS).
FT VARIANT 439 439 /FTId=VAR_005630.
FT VARIANT 810 810 P -> L.
FT VARIANT 810 810 /FTId=VAR_005631.
FT VARIANT 1313 1313 A -> V.
FT VARIANT 1355 1355 /FTId=VAR_005632.
FT VARIANT 1355 1355 A -> V (in dbSNP:15251).
FT CONFLICT 1312 1312 D -> G.
FT CONFLICT 1411 AA; 144312 MM; 3880203D985C2699 CRC64;
SQ SEQUENCE 1411 AA; 144312 MM; 3880203D985C2699 CRC64;

Query Match 64.7%; Score 70.5; DB 1; Length 1411;
Best Local Similarity 54.8%; Pred. No. 1.5;
Matches 17; Conservative 2; Mismatches 1; Indels 11; Gaps 1;

Qy 2 KKKKKKKKKKK-----KKKKKKK 21
Db 1375 KKKKKKKKKAKKASTKDSKSPKSKKKKKKK 1405

Search completed: January 30, 2004, 00:20:44
Job time : 5.73239 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 : Search time 21.6901 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-1
Perfect score: 109
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_fodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	100	91.7	55	4 Q8N6F0	Q8N6F0 homo sapien
2	100	91.7	80	10 Q8S7D3	Q8S7D3 cryza sativ
3	100	91.7	113	10 Q8LQP6	Q8LQP6 cryza sativ
4	100	91.7	128	3 Q9P529	Q9P529 neurospora
5	100	91.7	129	11 Q35807	Q35807 rattus norv
6	100	91.7	168	4 Q9H5V6	Q9H5V6 homo sapien
7	100	91.7	206	4 Q8I247	Q8I247 plasmodium
8	100	91.7	215	11 Q64075	Q64075 rattus sp.
9	100	91.7	260	10 Q9LGS29	Q9LGS29 arabidopsis
10	100	91.7	380	4 Q9NT34	Q9NT34 homo sapien
11	100	91.7	515	5 Q9SMR7	Q9SMR7 drosophila
12	100	91.7	517	10 Q9LXR2	Q9LXR2 arabidopsis
13	100	91.7	531	6 Q9SLV6	Q9SLV6 macaca fasc
14	100	91.7	720	4 Q9H6Q7	Q9H6Q7 homo sapien
15	100	91.7	791	5 Q8T2U7	Q8T2U7 dictyosteli
16	97	89.0	667	4 Q9HC48	Q9HC48 homo sapien

17	96	88.1	476	10 Q8RV44	Q8RV44 lycopersico
18	95	87.2	489	5 Q8T3H4	Q8T3H4 drosophila
19	95	87.2	658	11 Q8CGI8	Q8CGI8 mus musculu
20	92	84.4	4433	5 Q8LJ15	Q8LJ15 plasmodium
21	91	83.5	3351	5 Q8IBK4	Q8IBK4 plasmodium
22	90	82.6	39	5 Q8MSJ8	Q8MSJ8 drosophila
23	90	82.6	257	4 Q9H5Y3	Q9H5Y3 homo sapien
24	90	82.6	686	4 Q9NXX0	Q9NXX0 homo sapien
25	90	82.6	2820	5 Q8IM32	Q8IM32 plasmodium
26	89	81.7	83	11 Q9ER82	Q9ER82 mus musculu
27	89	81.7	1710	5 Q8I239	Q8I239 plasmodium
28	88	80.7	109	5 Q8I3G8	Q8I3G8 plasmodium
29	87	79.8	107	10 Q9LQF6	Q9LQF6 arabidopsis
30	87	79.8	476	5 Q8IFN3	Q8IFN3 plasmodium
31	86	78.9	102	11 Q8CSY8	Q8CSY8 mus musculu
32	86	78.9	741	5 Q8IPR6	Q8IPR6 drosophila
33	85	78.0	978	5 Q8IB83	Q8IB83 plasmodium
34	84	77.1	339	4 Q8IVS7	Q8IVS7 homo sapien
35	83	76.1	227	4 Q8TA86	Q8TA86 homo sapien
36	83	76.1	1337	5 Q8IC23	Q8IC23 plasmodium
37	82.5	75.7	710	5 Q8IDK7	Q8IDK7 plasmodium
38	82	75.2	943	5 Q8ILY0	Q8ILY0 plasmodium
39	81	74.3	253	11 Q8CFH9	Q8CFH9 mus musculu
40	81	74.3	547	4 Q8IV81	Q8IV81 homo sapien
41	81	74.3	1082	5 Q8ISY6	Q8ISY6 plasmodium
42	81	74.3	1313	5 Q8IKC0	Q8IKC0 plasmodium
43	81	74.3	3193	5 Q8IS90	Q8IS90 plasmodium
44	81	74.3	4550	5 Q77336	Q77336 plasmodium
45	80	73.4	213	11 P97762	P97762 mus musculu

ALIGNMENTS

RESULT 1

Q8N6F0 PRELIMINARY; PRT; 55 AA.

ID Q8N6F0

AC Q8N6F0:

DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Similar to LOC201361.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RA Strauberg R.;

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC030525; AAH0525.1; -

SQ SEQUENCE 55 AA; 7251 MW; 0906032B284006BA CRC64;

Query Match 91.7%; Score 100; DB 4; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KXXXXXXXXXXXXXXXXXXXX 21

Db 21 KXXXXXXXXXXXXXXXXXXXX 40

RESULT 2

ID Q8S7D3 PRELIMINARY; PRT; 80 AA.

AC Q8S7D3:

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Hypothetical 9.4 kDa protein.

GN OSJNBA0057L21.23.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaridaceae; Oryzae; Oryza.
 OK NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
 RA Ganabinger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
 RA Tselirin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pal G.,
 RA Vanhken S.E., Uteback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
 RA Salzberg S.L., White O., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBa0057121 genomic sequence.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC087599; AA179706.1; -.
 DR Gramene; Q857D3; -.
 KW Hypothetical protein.
 SQ SEQUENCE 80 AA; 9362 MW; 017C86313B21D8 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 80;
 Best Local Similarity 100.0%; Pred. No. 0.0001;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 48 KKKKKKKKKKKKKKKKKKK 67

RESULT 3

Q8LOP6 PRELIMINARY; PRT; 113 AA.

AC Q8LOP6; PRELIMINARY; PRT; 113 AA.
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE OJ117_G01.13 protein.
 GN OJ117_G01.13.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaridaceae; Oryzae; Oryza.
 OK NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsunoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
 clone:OJ117_G01.13";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003374; BAB93330.1; -.
 DR Gramene; Q8LOP6; -.
 SQ SEQUENCE 113 AA; 13660 MW; 597DB0BDEB2AA3EF CRC64;

Query Match 91.7%; Score 100; DB 10; Length 113;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 10 KKKKKKKKKKKKKKKKKKK 29

RESULT 4

Q9P529 PRELIMINARY; PRT; 128 AA.

AC Q9P529; PRELIMINARY; PRT; 128 AA.
 DT 01-DEC-2000 (TRENBLrel. 15, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical 15.2 kDa protein.
 GN B24H17.160.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.
 OK NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hehseisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL356815; CAB92638.2; -.
 KW Hypothetical protein.
 SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 91.7%; Score 100; DB 3; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00014;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 71 KKKKKKKKKKKKKKKKKKK 90

RESULT 5

O35807 PRELIMINARY; PRT; 129 AA.

AC O35807; PRELIMINARY; PRT; 129 AA.
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE MICROVASCULAR endothelial differentiation protein 2.
 GN MDG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OK NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISUB=Epilidymis;
 RX MEDLINE=98172708; Pubmed=9511718;
 RA Proels F., Loser B., Marx M.;
 RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
 species, during in vitro angiogenesis";
 RL Exp. Cell Res. 239:1-10(1998).
 DR EMBL; Y08769; CAA70022.1; -.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF00069; Kinase; 1.
 DR Prodom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR ATP-binding; Transferase
 SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 129;
 Best Local Similarity 100.0%; Pred. No. 0.00014;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 85 KKKKKKKKKKKKKKKKKKK 104

RESULT 6

Q9H5V6 PRELIMINARY; PRT; 168 AA.

AC Q9H5V6; PRELIMINARY; PRT; 168 AA.
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein FLJ22976 (Fragment).
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.


```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsashi T., Nishitani T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isigaki T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026629; BAB15513.1; -
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 140 KKKKKKKKKKKKKKKKKKK 159

RESULT 7
O81247 PRELIMINARY; PRT; 206 AA.
AC O81247;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
PF04775C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knighes A., Konfortov B., Kyes S., Larke N., Lawson D., Lemard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitz E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL031745; CAD49055.1; -
KW Hypothetical protein.
SQ SEQUENCE 206 AA; 25047 MW; 1192E49A3DC4523F CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 5; Length 206;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 185 KKKKKKKKKKKKKKKKKKK 204

RESULT 8
O64075 PRELIMINARY; PRT; 215 AA.
AC O64075;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

```

```

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Nucleoporin p62 homolog protein (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95151924; PubMed=7849178;
RA Wang Z.Q., Akmal K.M., Kim K.H.;
RT "An unusual nucleoporin-related messenger ribonucleic acid is present
RT in the germ cells of rat testis."
RL Biol. Reprod. 51:1022-1030(1994).
DR EMBL; S75997; AAB33384.1; -
KW Porin.
FT NON TER
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD8 CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 11; Length 215;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 35 KKKKKKKKKKKKKKKKKKK 54

RESULT 9
O9LG29 PRELIMINARY; PRT; 260 AA.
AC O9LG29;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone:FLD9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eusteroideae; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX SPRAIN=Columbia;
RA Nakamura Y.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002460; BAA97098.1; -
DR InterPro; IPR005819; Histonc_H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 10; Length 260;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 7 KKKKKKKKKKKKKKKKKKK 26

RESULT 10
O9NT34 PRELIMINARY; PRT; 380 AA.
AC O9NT34;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFP4341120.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RA Otterweiler B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117556; CAB70810.1; -
DR Genew; HGNC:15736; C17orf28.
KW Hypothetical protein.
FT NON TER 380
SQ SEQUENCE 380 AA; 42689 MW; 67F50DD101346AFB CRC64;

Query Match 91.7%; Score 100; DB 4; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 355 KKKKKKKKKKKKKKKKKKKKK 374

RESULT 11
ID Q8SWR7 PRELIMINARY; PRT; 515 AA.
AC Q8SWR7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE GH22607p (Fragment).
GN CG7180.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neuroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Aspayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Drenth D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guerin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY095518; AAM12251.1; -
DR FlyBase; FBgn0032673; CG7180.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SMO0194; PTPc; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolyase.
FT NON TER 515
SQ SEQUENCE 515 AA; 59080 MW; B2825B7BEA96195E CRC64;

Query Match 91.7%; Score 100; DB 5; Length 515;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 493 KKKKKKKKKKKKKKKKKKKKK 512

RESULT 12
ID Q9LXR2 PRELIMINARY; PRT; 517 AA.
AC Q9LXR2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

```

```

DE Hypothetical 59.7 kDa protein.
GN T20N10.250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Verzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALJ33032; CAB88307.1; -
DR InterPro; IPR001810; F-box.
DR InterPro; IPR006566; FBD.
DR Pfam; PF00646; F-box; 1.
DR SMART; SMO0579; FBD; 1.
DR PROSITE; PS0181; FBOX; 1.
DR PROSITE; PS50181; FBOX; 1.
KW Hypothetical protein.
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 444 KKKKKKKKKKKKKKKKKKKKK 463

RESULT 13
ID Q95LV6 PRELIMINARY; PRT; 531 AA.
AC Q95LV6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical 61.4 kDa protein (Fragment).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
CX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RA Hashimoto K., Osada N., Hida M., Kuehda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RA "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071085; BAB64479.1; -
KW Hypothetical protein.
FT NON TER 531
SQ SEQUENCE 531 AA; 61389 MW; B55996BAF5CDD60C CRC64;

Query Match 91.7%; Score 100; DB 6; Length 531;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 502 KKKKKKKKKKKKKKKKKKKKK 521

RESULT 14
ID Q9H6Q7 PRELIMINARY; PRT; 720 AA.
AC Q9H6Q7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

```

AC Q9H6Q7;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
 DE Hypothetical protein FLJ21979 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AK025632; BAB15196.1; -.
 KW Hypothetical protein.
 FT NON TER 720
 SQ SEQUENCE 720 AA; 84029 MW; A86586FEA953D0B CRC64;

Query Match 91.7%; Score 100; DB 4; Length 720;
 Best Local Similarity 100.0%; Pred. No. 0.00043;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
 DB 692 KKKKKKKKKKKKKKKKKKK 711

RESULT 15

ID O8T2U7 PRELIMINARY; PRT; 791 AA.
 AC O8T2U7;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Hypothetical 92.4 kDa protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC115574; AAL92183.1; -.
 DR InterPro; IPR005033; YEATS.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF03366; YEATS; 1.
 DR SMART; SM00355; ZNF_C2H2; 1.
 DR SMART; SM00355; ZNF_C2H2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 791 AA; 92375 MW; D66C8BDEC92352C CRC64;

Query Match 91.7%; Score 100; DB 5; Length 791;
 Best Local Similarity 100.0%; Pred. No. 0.00045;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
 DB 769 KKKKKKKKKKKKKKKKKKK 788

Search completed: January 30, 2004, 00:24:37
 Job time : 22.6901 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 10.6808 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-2
Perfect score: 109
Sequence: 1 CEAHAHAHAHAHAHAHAHAHAHAHA 25

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: dir1:*
2: dir2:*
3: dir3:*
4: dir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	70.6	265	2	S19113
2	72	66.1	165	2	B87702
3	71	65.1	205	2	S19114
4	70	64.2	183	2	S24960
5	70	64.2	401	2	A48423
6	69	63.3	97	2	S02376
7	69	63.3	1028	2	A56038
8	69	63.3	1213	2	S16356
9	68	63.3	1668	2	T13748
10	68	62.4	403	2	A81882
11	68	62.4	655	2	A29945
12	68	62.4	873	2	B53225
13	68	62.4	2023	2	T13154
14	67	61.5	314	2	UC5273
15	67	61.5	460	2	TJ3110
16	67	61.5	494	2	A42170
17	67	61.5	497	2	JC5076
18	67	61.5	604	2	A39369
19	67	61.5	606	2	S13670
20	67	61.5	627	2	T02610
21	67	61.5	671	2	C96534
22	67	61.5	1226	2	T24045
23	67	61.5	2715	2	T13049
24	66.5	60.1	543	2	B39369
25	65.5	60.1	273	2	T51010
26	65	59.6	85	1	A22592
27	65	59.6	333	2	A39065
28	65	59.6	475	2	A43915
29	65	59.6			

30	65	59.6	644	2	S39356	transcription fact
31	65	59.6	703	2	T48600	kinase-like protei
32	64	58.7	109	1	R487P1	acidic ribosomal p
33	64	58.7	392	2	B48423	homeotic protein e
34	64	58.7	1065	2	T13230	dachshund isoform
35	64	58.7	1072	2	T13232	dachshund protein
36	64	58.7	1074	2	T13229	dachshund protein
37	64	58.7	1081	2	T13231	abdominal segment
38	64	58.7	1533	2	A46221	female sterile hom
39	64	58.7	2038	2	A43742	50S ribosomal prot
40	63	57.8	179	2	AF2908	polyomavirus enhan
41	63	57.8	179	2	F97683	asparaginyl-tna s
42	63	57.8	513	2	A48233	conserved hypochne
43	63	57.8	568	2	T39675	GTP-binding regula
44	63	57.8	581	2	E75383	
45	63	57.8	846	2	S52418	

ALIGNMENTS

RESULT 1
S19113
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
C:Species: Chlamydomonas reinhardtii
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S19113; S14466
R:Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A:Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A:Reference number: S19113; MUID:92119224; PMID:1731966
A:Accession: S19113
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <MAX>
A:Cross-references: EMBL:X17208; NID:g18136; PIDN:CA35080.1; PID:g18137
C:Genetics:
A:Gene: cgcr-4

Query Match 70.6% Score 77; DB 2; Length 265;
Best Local Similarity 82.6% Pred. No. 0.25;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHAHAHAHAHAHAHAHAHA 25
DB 154 AHAHAHAHAHAHAHAHAHA 176

RESULT 2
B87702
ribosomal protein S16 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87702
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J.; B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, n, J.; Smolova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87702
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-165 <STO>
A:Cross-references: GB:AE005673; NID:g13425408; PIDN:AAK25614.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3652

Query Match 66.1% Score 72; DB 2; Length 165;
Best Local Similarity 75.0% Pred. No. 0.56;
Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EHAHAHAHAHAHAHAHAHA 25

DB 115 QAEADAKAAAEKAAAEAAAAA 138

RESULT 3

S19114
cscr-1 protein - Chlamydomonas reinhardtii (fragment)
C.Species: Chlamydomonas reinhardtii
C.Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 03-Nov-2000
C.Accession: S19114
R.Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A.Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A.Reference number: S19113; MUID:92119224; PMID:1731966
A.Accession: S19114
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-205 <MAX>
A.Cross-references: EMBL:X17207
C.Superfamily: phage lambda hypothetical protein 401

Query Match 65.1%; Score 71; DB 2; Length 205;
Best Local Similarity 75.0%; Pred. No. 0.82;
Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEAAAAAEAAAAAEAAAAA 25
DB 49 EEAADAAQAAAAAEAAAAAEAAAAA 72

RESULT 4

S24960
gene C98 protein - rape
C.Species: Brassica napus (rape)
C.Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C.Accession: S24960
R.Roberts, M.R.; Hodge, R.; Sorensen, A.; Ross, J.; Murphy, D.J.; Draper, R.; Scott, R.
submitted to the EMBL Data Library, July 1992
A.Description: A new class of Brassica napus oleosin genes specific to the male gametoph
A.Reference number: S24960
A.Accession: S24960
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-183 <ROB>
A.Cross-references: EMBL:X67142; NID:g17792; PID:g17793

Query Match 64.2%; Score 70; DB 2; Length 183;
Best Local Similarity 78.3%; Pred. No. 0.95;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAAEAAAAAEAAAAA 25
DB 151 AAPAAEPAPAAEAAPAAEAAPAA 173

RESULT 5

A48423
engrafted homeodomain-containing protein En-1 - mouse
N/Alternate names: homeotic protein En-1
C.Species: Mus musculus (house mouse)
C.Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C.Accession: A48423; S13009; A26629; A24778
R.Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.L
Dev. Genet. 13, 345-358, 1992
A.Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene
A.Reference number: A48423; MUID:93185339; PMID:1363401
A.Accession: A48423
A.Status: preliminary; not compared with conceptual translation
A.Molecule type: mRNA
A.Residues: 1-401 <LOG>
A.Experimental source: CD-1, embryo
A.Note: sequence extracted from NCBI backbone (NCBIP:126620)
R.Holland, P.W.H.; Williams, N.A.

FEBS Lett. 277, 250-252, 1990
A.Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution.
A.Reference number: S13009; MUID:91099509; PMID:1980115
A.Accession: S13009
A.Status: preliminary
A.Molecule type: nucleic acid
A.Residues: 321-380 <HOL>
R.Joyner, A.L.; Martin, G.R.
Genes Dev. 1, 29-38, 1987

A.Title: En-1 and Bn-2, two mouse genes with sequence homology to the Drosophila engrailed
A.Reference number: A91620; MUID:88112776; PMID:2892757
A.Accession: A26629
A.Molecule type: DNA; mRNA
A.Residues: 278-401 <JOY>
A.Cross-references: GB:Y00201; GB:M11987; NID:g49587; PID:CA68361.1; PID:g669105
R.Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.
Cell 43, 29-37, 1985
A.Title: Expression during embryogenesis of a mouse gene with sequence homology to the D
A.Reference number: A24778; MUID:86079501; PMID:2416459
A.Accession: A24778
A.Molecule type: DNA
A.Residues: 311-401 <JO2>
C.Genetics:
A.Gene: en.1
A.Map position: 1

C.Superfamily: unassigned homeobox proteins; homeobox homology
C.Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:313-369/Domain: homeobox homology <HOX>

Query Match 64.2%; Score 70; DB 2; Length 401;
Best Local Similarity 78.3%; Pred. No. 1.7;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAAEAAAAAEAAAAA 25
DB 207 AAAAAAATAAAAAAATAAAAAA 229

RESULT 6

S02376
antifreeze protein precursor - yellowtail flounder
C.Species: Limanda ferruginea (yellowtail flounder)
C.Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 24-Oct-2000
C.Accession: S02376
R.Scott, G.K.; Davies, P.L.; Shears, M.A.; Fletcher, G.L.
Eur. J. Biochem. 168, 629-633, 1987
A.Title: Structural variations in the alanine-rich antifreeze proteins of the Pleuronecti
A.Reference number: S02376; MUID:88029483; PMID:3665937
A.Accession: S02376
A.Molecule type: mRNA
A.Residues: 1-97 <SCO>
A.Cross-references: EMBL:X06356; NID:g64041; PID:CAA29655.1; PID:g64042
A>Note: part of this sequence, including the amino end of the mature protein, was confirm
C.Superfamily: antifreeze protein
C.Keywords: antifreeze
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-48/Domain: propeptide #status predicted <PRO>
F:49-96/Product: antifreeze protein #status predicted <MAT>

Query Match 63.3%; Score 69; DB 2; Length 97;
Best Local Similarity 70.8%; Pred. No. 0.75;
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEAAAAAEAAAAAEAAAAA 25
DB 53 DAAAAAATAAAAAAATAAAAAA 76

RESULT 7

A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C.Species: Drosophila melanogaster
C.Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000

C/Accession: A56038
R/Gartinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A/Title: Multiple products from the shavenbaby-ovo gene region of *Drosophila melanogaster*
A/Reference number: A56038; MUID:95021209; PMID:7935398
A/Accession: A56038
A/Status: Preliminary
A/Molecule type: mRNA
A/Residues: 1-1028 <GAR>
A/Cross-references: GB:U11383; NID:9520526; PIDN:AA60216.1; PID:9520527
C/Genetics:
A/Gene: ovo
A/Cross-references: FlyBase:FBgn0003028

Query Match 63.3%; Score 69; DB 2; Length 1028;
Best Local Similarity 85.7%; Pred. No. 4.2;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAAAAAAAAAAAAAA 23
Db 497 AAAAAAAAAAAAAAAAAAAAAA 517

RESULT 8
S16356
ovo protein - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C/Accession: S16356
R/Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2259-2266, 1991
A/Title: The ovo gene of *Drosophila* encodes a zinc finger protein required for female sex
A/Reference number: S16356; MUID:91293102; PMID:1712294
A/Accession: S16356
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-1213 <MEV>
A/Cross-references: EMBL:X59772
C/Genetics:
A/Gene: FlyBase:ovo
A/Cross-references: FlyBase:FBgn0003028
A/Interons: 931/3; 1152/3

Query Match 63.3%; Score 69; DB 2; Length 1213;
Best Local Similarity 85.7%; Pred. No. 4.7;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAAAAAAAAAAAAAA 23
Db 860 AAAAAAAAAAAAAAAAAAAAAA 880

RESULT 9
T13748
sex comb protein - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C/Accession: T13748
R/Sinclair, D.A.R.; Milne, T.A.; Hodgson, J.W.; Shellard, J.; Salinas, C.A.; Kyba, M.; R
Development 125, 1207-1216, 1998
A/Title: The Additional sex combs gene of *Drosophila* encodes a chromatin protein that b
A/Reference number: Z17750; MUID:98146384; PMID:9477319
A/Accession: T13748
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1668 <SIN>
A/Cross-references: EMBL:AJ001164; NID:93292938; PIDN:CAA04566.1; PID:93292933
C/Genetics:
A/Cross-references: FlyBase:FBgn0000142
C/Function:
A/Description: involved in repression of homeotic loci

Query Match 63.3%; Score 69; DB 2; Length 1668;

Best Local Similarity 78.3%; Pred. No. 6;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EAAAAAAAAAAAAAAAAAAAAA 24
Db 128 KAAAAAAAAAAAAAAAAAAAAA 150

RESULT 10
A81882
probable dihydroilipoamide S-succinyltransferase (EC 2.3.1.61) E2 component NMA1150 [impor
C/Species: *Neisseria meningitidis*
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C/Accession: A81882
R/Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A/Reference number: A81775; MUID:20222556; PMID:10761919
A/Accession: A81882
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-403 <PAR>
A/Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CA884412.1; PID:G737984;
A/Experimental source: serogroup A, strain Z2491
C/Genetics:
A/Gene: sucP, NMA1150
C/Superfamily: dihydroilipoamide acetyltransferase; lipoyl/biotin-binding homology
C/Keywords: acyltransferase; coenzyme A

Query Match 62.4%; Score 68; DB 2; Length 403;
Best Local Similarity 75.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 EAAAAAAAAAAAAAAAAAAAAA 25
Db 85 EAPAAATTAATAAPAAAPAA 108

RESULT 11
A29945
neurogenesis regulatory protein - fruit fly (*Drosophila melanogaster*) (fragment)
N/Alternate names: single-minded gene protein
C/Species: *Drosophila melanogaster*
C/Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C/Accession: A29945
R/Crews, S.T.; Thomas, J.B.; Goodman, C.S.
Cell 52, 143-151, 1988
A/Title: The *Drosophila* single-minded gene encodes a nuclear protein with sequence simil
A/Reference number: A29945; MUID:88151023; PMID:3345560
A/Accession: A29945
A/Molecule type: mRNA
A/Residues: 1-655 <CRE>
A/Cross-references: GB:ML9020; NID:G158464; PID:G158465
C/Genetics:
A/Gene: sim
A/Cross-references: FlyBase:FBgn0004666
C/Keywords: DNA binding; transcription regulation

Query Match 62.4%; Score 68; DB 2; Length 655;
Best Local Similarity 66.7%; Pred. No. 3.8;
Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EAAAAAAAAAAAAAAAAAAAAA 25
Db 366 QAAQAQAQAQAQAQAQAQAQA 389

RESULT 12
B53225
ecdysone-induced protein E74A - fruit fly (*Drosophila virilis*)
C/Species: *Drosophila virilis*
C/Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Feb-1997

```

C/Accession: B53225
R:Jones, C.W.; Dalton, M.W.; Tomlin, L.H.
Genetics 127, 535-543, 1991
A>Title: Interspecific comparisons of the structure and regulation of the Drosophila ecd
A/Reference number: A53225; MUID:91200627; PMID:2016053
A/Accession: B53225
A>Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-873 <JON>
A/Cross-references: GB:X59493
C/Genetics:
A/Gene: FlyBase:Divit/Bip74EF
A/Cross-references: FlyBase:FBgn0013076
C/Superfamily: ets DNA-binding domain homology
F:779-855/Domain: ets DNA-binding domain homology <ETS>

Query Match      62.4%; Score 68; DB 2; Length 873;
Best Local Similarity 73.9%; Pred. No. 4.7;
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      3 AAAAAAAAAAAAAAAAAAAAAA 25
Db      501 AAAAAAAAAAATGSAAAAAAAAA 523

RESULT 13
T33154
polycomb protein enhancer - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C/Accession: T33154
R:Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.
Development 125, 4055-4066, 1998
A>Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein conserv
A/Reference number: Z17611; MUID:98407961; PMID:9735366
A/Accession: T33154
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2023 <STR>
A/Cross-references: EMBL:AF079764; NID:93757889; PID:93757890; PIDN:AA64271.1
A/Experimental source: imaginal disc
C/Genetics:
A/Gene: B(Pc)
A/Cross-references: FlyBase:FBgn0000581
A/Map position: 2

Query Match      62.4%; Score 68; DB 2; Length 2023;
Best Local Similarity 60.0%; Pred. No. 8.7;
Matches 18; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

QY      1 CE-----AAAAAAAAAAAAAAAAAAAA 24
Db      810 CEDQVASTSAAAAAAAAAAAAAAAAAAS 839

RESULT 14
JCS273
paired type homeobox protein, NBP - human
C/Species: Homo sapiens (man)
C/Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 24-Sep-1999
C/Accession: JCS273
R:Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.
DNA Res. 3, 311-320, 1996
A>Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific
A/Reference number: JCS272; MUID:97191543; PMID:9039501
A/Contents: neuroblastoma cell
A/Accession: JCS273
A>Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-314 <YOK>
A/Cross-references: DBJ:D8244; NID:9184337; PIDN:BA11555.1; PID:d1012222; PID:918413
C/Comment: This protein is a transcriptional repressor involved in regulating gene expre
C/Superfamily: unassigned homeobox proteins; homeobox homology

```

```

C/Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:99-155/Domain: homeobox homology <HOX>

Query Match      61.5%; Score 67; DB 2; Length 314;
Best Local Similarity 78.3%; Pred. No. 2.8;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      3 AAAAAAAAAAAAAAAAAAAAAA 25
Db      244 AAAAAAAAAAAAAAAAAAGGLAA 266

RESULT 15
T33110
hypothetical protein C18H7.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C/Accession: T33110
R:Tin-Wollam, A.; Fronick, W.
submitted to the EMBL Data Library, May 1998
A/Description: The sequence of C. elegans cosmid C18H7.
A/Reference number: Z21284
A/Accession: T33110
A/Molecule type: DNA
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Residues: 1-460 <TIN>
A/Cross-references: EMBL:AF067607; PIDN:AACT7641.1; GSPDB:GN00022; CESP:C18H7.3
A/Experimental source: strain Bristol N2; clone C18H7
C/Genetics:
A/Gene: CESP:C18H7.3
A/Map position: 4
A/Introns: 84/1
C/Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match      61.5%; Score 67; DB 2; Length 460;
Best Local Similarity 62.5%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 4; Indels 8; Gaps 1;

QY      2 EAAAAAAAAA-----AAAAAAAAAA 25
Db      399 EAAAPAPAPAPAGAGGAPAPAPAPAAAA 430

```

Search completed: January 30, 2004, 00:26:21
 Job time : 11.6808 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 5.6338 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-2
Perfect score: 109
Sequence: 1 CEAATAAEEAATAAEEAATAAEEAATAA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	72.5	332	1	PAVA_HUMAN
2	75.5	63.3	518	1	TPM4_DROME
3	72	66.1	165	1	RS16_CAUCR
4	70	64.2	183	1	OLEC3_BRAVA
5	70	64.2	401	1	HME1_MOUSE
6	69	63.3	97	1	ANP_LIMFE
7	69	63.3	1028	1	OVO_DROME
8	69	63.3	1669	1	ASX_DROME
9	68	62.4	697	1	SIW_DROME
10	68	62.4	1073	1	HR38_DROME
11	67	61.5	314	1	PKXB_HUMAN
12	67	61.5	314	1	PKXB_MOUSE
13	67	61.5	477	1	MA2_HUMAN
14	67	61.5	606	1	HMID_DROAN
15	66	60.6	386	1	HXAD_MOUSE
16	66	60.6	388	1	HXAD_HUMAN
17	65	59.6	85	1	ANP4_PSEAM
18	65	59.6	91	1	ANPX_PSEAM
19	65	59.6	276	1	SK21_HUMAN
20	65	59.6	280	1	SK21_CHICK
21	65	59.6	475	1	EVX2_MOUSE
22	65	59.6	476	1	EVX2_HUMAN
23	65	59.6	644	1	BTJ_DROME
24	64	58.7	109	1	RLAI_TRYCR
25	64	58.7	392	1	HME1_HUMAN
26	64	58.7	1533	1	PUM_DROME
27	64	58.7	2038	1	FSH_DROME
28	63	57.8	179	1	RL19_AGRIS
29	63	57.8	376	1	FXE1_HUMAN
30	63	57.8	521	1	RUN2_HUMAN
31	63	57.8	562	1	ARX_HUMAN
32	63	57.8	590	1	HMAA_DROME
33	63	57.8	607	1	RUN2_MOUSE

ALIGNMENTS

34	63	57.8	1095	1	PIPA_DROME	P13217 drosophila
35	62	56.9	364	1	NGK1_MESAU	O60554 mesocricetu
36	62	56.9	365	1	NGK1_MOUSE	O99949 mus musculu
37	62	56.9	365	1	NGK1_RAT	O35762 rattus norv
38	62	56.9	1355	1	SALM_DROME	P35770 drosophila
39	61	56.0	31	1	ANP3_PAGOBO	P02732 pagothenia
40	61	56.0	91	1	ANP3_PSEAM	P23699 pseudopleur
41	61	56.0	153	1	RS16_BIFLO	O89791 bifidobacte
42	61	56.0	308	1	ABE1_DROME	P18413 drosophila
43	61	56.0	376	1	FXL2_HUMAN	P58012 homo sapien
44	61	56.0	421	1	PO41_MOUSE	P17208 mus musculu
45	61	56.0	423	1	PO41_HUMAN	Q01851 homo sapien

RESULT 1
FA9A_HUMAN STANDARD; PRT; 332 AA.

AC O81ZIU;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein FAM9A.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX MEDLINE=22202142; PubMed=12213195;
RA Martinez-garay I., Jablonka S., Sutajova M., Steuernagel P., Gal A.,
RA Kutsche K.;
RT "A new gene family (FAM9) of low-copy repeats in Xp22.3 expressed
RT exclusively in testis: implications for recombinations in this
RT region.";
RL Genomics 80:259-267(2002).

CC - SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC - TISSUE SPECIFICITY: Expressed exclusively in testis.
CC - SIMILARITY: Belongs to the FAM9 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; AF494343; AAN07162.1; -
DR Genew; HGNC:18403; FAM9A.
KW Nuclear protein.
FT 180 275 GLU-RICH.
FT DOMAIN 194 214 POLY-ALA.
FT 252 258 POLY-GLY.
SQ SEQUENCE 332 AA; 37339 MW; 92F22EC36038229C CRC64;

Query Match 72.5%; Score 79; DB 1; Length 332;
Best Local Similarity 83.3%; Pred. No. 0.066;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAEEAATAAEEAATAAEEAATAA 25
DB 190 EAEAEAAATAAEEAATAAEEAATAA 213

RESULT 2
TPM4_DROME STANDARD; PRT; 518 AA.
AC P49455; P49456; Q24425; Q24426;
DT 01-FEB-1996 (Rel. 33, Created)

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AB006023; AAK25614.1; --
 DR PIR; B87702; B87702.
 DR HSSP; P80379; 1EMW.
 DR TIGR; CC3652; --
 DR HAMAP; MF_00385; -- 1.
 DR InterPro; IPR000307; Ribosomal_S16.
 DR Pfam; PF00886; Ribosomal_S16; 1.
 DR Prodom; PD003791; Ribosomal_S16; 1.
 DR TIGRFAMs; TIGR00002; S16; 1.
 DR PROSITE; PS00732; RIBOSOMAL_S16; 1.
 DR Ribosomal protein, Complete proteome.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 165 AA; 17605 MW; ED46FC2798C5BE1C CRC64;

Query Match 66.1%; Score 72; DB 1; Length 165;
 Best Local Similarity 75.0%; Pred. No. 0.19;
 Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 115 QAEADAKAAAEKXAAAAAAAA 138

RESULT 4
 OLEC_BRANA STANDARD; PRT; 183 AA.
 AC P29526; 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Oleosin C98 (Fragment).
 GN C98
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=anther;
 RX MEDLINE=93386188; PubMed=8374615;
 RA Roberts M.R., Hodge R., Rose J.H.E., Sorensen A., Murphy D.J.,
 RA Draper J., Scott R.;
 RT "Characterization of a new class of oleosins suggests a male
 gametocyte-specific lipid storage pathway";
 RL Plant J. 3:629-636(1993).
 RT FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY
 DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE
 OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOLECULES
 OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC
 LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.
 CC MONOLAYER LIPID/WATER INTERFACE.
 CC SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A
 CC -1- TISSUE SPECIFICITY: SPECIFIC TO THE MALE GAMETOPHYTE.
 CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X67142; CAA47623.1; --
 DR PIR; S24960; S24960.
 DR InterPro; IPR000136; Oleosin.

DR Pfam; PF01277; Oleosin; 1.
 DR PROSITE; PS00811; OLEOSINS; 1.
 KW Seed; Oil body; Multigene family.
 FT NON_TER 1
 FT DOMAIN <1 23 POLAR.
 FT DOMAIN 24 95 HYDROPHOBIC.
 SQ SEQUENCE 183 AA; 18149 MW; 198A5D3BD6F3045A CRC64;

Query Match 64.2%; Score 70; DB 1; Length 183;
 Best Local Similarity 78.3%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
 DB 151 AAPAAPAPAPAPAPAPAPAPAA 173

RESULT 5
 HME1_MOUSE STANDARD; PRT; 401 AA.
 AC P09065; 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homeobox protein engrailed-1 (Mo-Eh-1).
 GN EN1 OR EN-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93185339; PubMed=1363401;
 RA Logan C., Hanks M.C., Noble-Topham S., Nallathattan D.,
 RA Provart N.J., Joyner A.L.;
 RT "Cloning and sequence comparison of the mouse, human, and chicken
 RT engrailed genes reveal potential functional domains and regulatory
 RT regions";
 RL Dev. Genet. 13:345-358(1992).
 RN [2]
 RP SEQUENCE OF 278-401 FROM N.A.
 RX MEDLINE=88112776; PubMed=2892757;
 RA Joyner A.L., Martin G.R.;
 RT "En-1 and En-2, two mouse genes with sequence homology to the
 RT Drosophila engrailed gene: expression during embryogenesis";
 RL Genes Dev. 1:29-38(1987).
 RN [3]
 RP SEQUENCE OF 298-401 FROM N.A.
 RX MEDLINE=86079501; PubMed=2416459;
 RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
 RT "Expression during embryogenesis of a mouse gene with sequence
 RT homology to the Drosophila engrailed gene";
 RL Cell 43:29-37(1985).
 RN [4]
 RP SEQUENCE OF 321-380 FROM N.A.
 RX MEDLINE=91095509; PubMed=1980115;
 RA Holland P.W.H., Williams N.A.;
 RT "Conservation of engrailed-like homeobox sequences during vertebrate
 RT evolution";
 RL FEBS Lett. 277:250-252(1990).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ENGRAILED HOMEBOX FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U12703; AAA03660.2; --

```

DR EMBL: Y00201; CAA68361.1; -.
DR PIR: A48423; A48423.
DR HSSP: P02836; 3HSD.
DR TRANSFAC: T02016; -.
DR MGD; MGI:95389; Enl.
DR InterPro; IPR000747; Engrailed.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_Lambrprepressr.
DR Pfam; PF00046; homeobox_1.
DR PRINTS; PR00024; ENGRAILED.
DR PRINTS; PR00031; HTHREPRESSR.
DR Prodom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS00033; ENGRAILED; 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
KW DOMAIN 52 87 PRO-RICH.
FT DOMAIN 73 87 POLY-ALA.
FT DOMAIN 207 228 POLY-ALA.
FT DNA_BIND 312 371 HOMEBOX.
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152PFAE CRC64;

Query Match 64.2%; Score 70; DB 1; Length 401;
Best Local Similarity 78.3%; Pred. No. 0.62;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 EMBL: Y00201; CAA68361.1; -.
DB 207 EMBL: Y00201; CAA68361.1; -.

RESULT 6
ANP_LIMFE STANDARD; PRT; 97 AA.
ID ANP_LIMFE STANDARD; PRT; 97 AA.
AC P09031;
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Antifreeze protein precursor (AFP).
OS Limnoria ferruginea (Yellowtail flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes;
OC Pleuronectidae; Pleuronectidae; Limanda.
OX NCBI_Taxid=8258;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=88029483; PubMed=3665937;
RA Scott G.K.; Davies P.L.; Shears M.A.; Fletcher G.L.;
RT "Structural variations in the alanine-rich antifreeze proteins of the
RT pleuronectinae."
RL Eur. J. Biochem. 168:629-633(1987).
CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X06356; CAA29655.1; -.
DR PIR: S02376; S02376.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE.
KW Antifreeze protein; Repeat; Signal.
FT SIGNAL 1 23
FT PROPEP 24 48 REMOVED BY A DIPEPTIDYLPEPTIDASE

```

```

FT CHAIN 49 97 (PROBABLE).
SQ SEQUENCE 97 AA; 8865 MW; 62AD582DF8E459B6 CRC64;

Query Match 63.3%; Score 69; DB 1; Length 97;
Best Local Similarity 70.8%; Pred. No. 0.26;
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EMBL: Y00201; CAA68361.1; -.
DB 53 EMBL: Y00201; CAA68361.1; -.

RESULT 7
OVO_DROME STANDARD; PRT; 1028 AA.
ID OVO_DROME STANDARD; PRT; 1028 AA.
AC P51521; Q9XZU4;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ovo protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Ephydroidea; Ephydroidea; Diptera; Brachycera; Muscomorpha;
OC NCB1_Taxid=7227;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Ovary;
RC MEDLINE=95021209; PubMed=7935398;
RA Garfinkel M.D.; Wang J.; Liang Y.; Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
RT melanogaster: relationship to genetic complexity."
RL Mol. Cell. Biol. 14:6809-6818(1994).
[2]
SEQUENCE FROM N.A.
RP STRAIN=Oregon-R;
RC MEDLINE=91293102; PubMed=1712294;
RA Wevel-Minto M.T.M.; Terracol R.; Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required
RT for female germ line development."
RL EMBL: J. 10:2259-2266(1991).
CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U11383; AAB60216.1; -.
DR EMBL: X59772; CAB36921.1; ALT_SEQ.
DR PIR: A56038; A56038.
DR HSSP; P07248; ZADR.
DR TRANSFAC; T00669; -.
DR FLYBase; FBgn0003028; ovo.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; ZF-C2H2_3.
DR SMART; SM00355; ZNF_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 3.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.

```

```

FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 114 122 POLY-HIS.
FT DOMAIN 133 159 POLY-ASN.
FT DOMAIN 336 359 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 539 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN FING 874 896 C2H2-TYPE 1.
FT ZN FING 902 924 C2H2-TYPE 2.
FT ZN FING 930 953 C2H2-TYPE 3.
FT ZN FING 969 992 C2H2-TYPE 4.
FT ZN FING 969 992 A -> R (IN REF. 2).
FT ZN FING 969 992 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D7068B82BC06F77 CRC64;

Query Match
Best Local Similarity 85.7%; Score 69; DB 1; Length 1028;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 23
DB 497 AAAAAAAAAAAAAAAAAAAAAA 517

RESULT 8
ASX DROME STANDARD; PRT; 1669 AA.
AC 09V727; 076930;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Polycarb protein Asx (Additional sex combs).
GN ASX OR CG8787.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI TaxID=7227;
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND DEVELOPMENTAL
STAGE.
RC TISSUE=Imaginal disks;
RX MEDLINE=9814684; PubMed=9477319;
RA Similair D.A.R., Milne T.A., Hodgson J.W., Shellard J., Salinas C.A.,
RA Kyba M., Randazzo P., Brock H.W.;
RA "The Additional sex combs gene of Drosophila encodes a chromatin
protein that binds to shared and unique polycarb group sites on
polytene chromosomes.";
RT Development 125:1207-1216 (1998).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Decher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.D., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li Z., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazzolo M., Pletman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RP INTERACTION WITH TAN.
RX MEDLINE=21290825; PubMed=11397012;
RA Dietrich B.H., Moore J., Kyba M., dosSantos G., McClokey F.,
RA Milne T.A., Brock H.W., Krause H.M.;
RT "Tantalus", a novel ASX-interacting protein with tissue-specific
functions.";
RL Dev. Biol. 234:441-453 (2001).
CC -1- FUNCTION: Atypical Polycarb group protein, which may be involved
in both Polycarb group (PCG) and trithorax group (trxg) complexes.
PCG and trxg proteins act by forming multiprotein complexes, which
are respectively required to maintain the transcriptionally
repressive and transcriptionally active state of homeotic genes
throughout development. PCG and trxg protein complexes are not
required to initiate repression and activation, but to maintain it
during later stages of development. Both complexes probably act
via methylation of histones, rendering chromatin heritably changed
in its expressibility.
CC -1- SUBUNIT: Interacts with Tan.
CC -1- SUBCELLULAR LOCATION: Nuclear; associated with chromatin.
Colocalizes with many PCG sites on polytene chromosomes. It also
associates with many unique sites on polytene chromosomes.
CC -1- TISSUE SPECIFICITY: Highly expressed in nurse cells and deposited
in oocytes late in oogenesis. Ubiquitous in early embryos. Late
embryos show higher levels in CNS and neuroectoderm.
CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
Early embryos have high levels of expression, this drops off and
zygotic expression begins at 3-6 hour embryos. Expression levels
are low in larvae and medium in pupae and adults.
CC -1- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXLL) motifs, which may
be required for an association with nuclear receptors (By
similarity).
CC -1- SIMILARITY: Belongs to the Asx family.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to
frameshifts in positions 608 and 719.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
CC or send an email to license@isb-stb.ch).

CC -----
CC EMBL, AJ001164; CA04568.1; ALT_FRAME.
CC FLYBASE, AEO03814; AAF58239.1; -.
CC DR PROSITE, PS01359; ZF_PHD_1; FALSE_NEG.
CC DR PROSITE, PS50016; ZF_PHD_2; FALSE_NEG.
CC KW Transcription regulation; Repressor; Nuclear protein; Zinc;
CC Metal-binding; Zinc-finger; Repeat; Developmental protein.
CC PHD-TYPE (ATYPICAL).
CC FT ZN_FING 1669 1669
CC FT DOMAIN 8 12
CC FT DOMAIN 122 126 POLY-GLN.
CC FT DOMAIN 129 152 ALA-RICH.
CC FT DOMAIN 638 715 SER-RICH.
CC FT DOMAIN 747 751 POLY-GLN.
CC FT DOMAIN 862 1202 GLN-RICH.
CC FT DOMAIN 1287 1290 POLY-THR.
CC FT DOMAIN 1520 1524 POLY-HIS.
CC FT SITE 1527 1536 POLY-GLN.
CC FT SITE 224 228 LKKLL MOTIF 1.
CC FT SITE 244 248 LKKLL MOTIF 2.
CC FT CONFLICT 14 15 SQ -> CE (IN REF. 1).
CC FT CONFLICT 187 187 K -> N (IN REF. 1).
CC FT CONFLICT 1253 1253 S -> T (IN REF. 1).
CC FT CONFLICT 1520 1520 MISSING (IN REF. 1).
CC SQ SEQUENCE 1669 AA; 179641 MW; F65D87398D6D321 CRC64;

Query Match 63.3%; Score 69; DB 1; Length 1669;
Best Local Similarity 78.3%; Pred. No. 2.4;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 24
Db 128 KAAAAAAAAAAAAAAAAAAAAA 150

RESULT 9
SIM_DROME STANDARD; PRT; 697 AA.
ID SIM_DROME 096521; Q8MOJ7; Q9VEZ3;
AC P05709; 096521; Q8MOJ7; Q9VEZ3;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Single-minded protein.
DE SIM OR CG7771.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCB1_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99054545; PubMed=9840810;
RA Kaasli Y., Stahl S., Crews S.;
RT "Specification of the Drosophila CNS midline cell lineage: direct
RT control of single-minded transcription by dorsal/ventral patterning
RT genes";
RL Gene Expr. 7:171-189(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake J.R., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKus G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Fabios B., Delcher A., Deng Z., Nays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Noy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster";
RA Science 287:2185-2195(2000).
RN [3]
RP REVISIONS.
RX STRAIN=Berkley;
MEDLINE=22426069; PubMed=12537572;
RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Belletcourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley; TISSUE=Embryo;
MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champ M.,
RA George R.A., Guarni H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [5]
RP SEQUENCE OF 25-42 FROM N.A., AND SIMILARITY TO HLH PROTEINS.
RX MEDLINE=92103681; PubMed=1760843;
RA Nabu J.R., Lewis J.O., Wharton K.A., Jr., Crews S.T.;
RT "The Drosophila single-minded gene encodes a helix-loop-helix protein
RT that acts as a master regulator of CNS midline development";
RL Cell 67:1157-1167(1991).
RN [6]
RP SEQUENCE OF 43-697 FROM N.A.
RX MEDLINE=88151023; PubMed=3345560;
RA Crews S.T., Thomas J.B., Goodman C.S.;
RT "The Drosophila single-minded gene encodes a nuclear protein with
RT sequence similarity to the per gene product";
RL Cell 52:143-151(1988).
RN [7]
RP FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER
RP DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN
RP THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING
RP RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.
CC -1- SUBUNIT: Efficient DNA binding requires dimerization with another
CC BHLH protein.

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- POLYMORPHISM: Berkeley strain has 11 A-A-Q repeats.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS.
 CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AF071934; AAC64519.1; ALT_SEQ.
 CC EMBL, AE003698; AAF54902.3; -.
 CC EMBL, AY129457; AAM76199.1; -.
 CC EMBL, M19020; AAB28900.1; -.
 CC PIR, A29945; A29945.
 CC TRANSFAC, T00750; -.
 CC FLYBase, FBgn0004666; sim.
 CC DR GO, GO:0005634; C:nucleus, IEP.
 CC DR GO, GO:0003702; F:RNA polymerase II transcription factor acti., NMS.
 CC DR GO, GO:0006355; P:regulation of transcription, DNA-dependent, NMS.
 CC DR GO, GO:0007418; P:ventral midline development, IMP.
 CC DR InterPro, IPR001092; HLH_baslc.
 CC DR InterPro, IPR001067; Nuc_translocat.
 CC DR InterPro, IPR001610; PAC.
 CC DR InterPro, IPR000014; PAS_domain.
 CC DR Pfam, PF00010; HLH; 1.
 CC DR Pfam, PF00785; PAC; 1.
 CC DR Pfam, PF00989; PAS; 1.
 CC DR PRINTS, PR00785; NCTRNSLOCATR.
 CC DR SMART, SMO0353; HLH; 1.
 CC DR SMART, SMO0086; PAC; 1.
 CC DR SMART, SMO0091; PAS; 2.
 CC DR TIGRFAMs, TIGR00229; sensory_box, 2.
 CC DR PROSITE, PS00038; HLH_1; 1.
 CC DR PROSITE, PS00888; HLH_2; 1.
 CC DR PROSITE, PS0112; PAS; 2.
 CC KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;
 CC Transcription regulation; DNA-binding.
 CC FT DNA_BIND 21 37
 CC FT DOMAIN 38 78 BASIC DOMAIN.
 CC FT DOMAIN 100 172 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 CC FT DOMAIN 266 336 PAS 1.
 CC FT DOMAIN 406 446 PAS 2.
 CC FT DOMAIN 553 672 SER-RICH.
 CC FT DOMAIN 673 693 GLN/HIS-RICH.
 CC FT VARIANT 406 414 MISSING (IN STRAIN BERKELEY).
 CC FT CONFLICT 151 151 I -> Y (IN REF. 1).
 CC SQ SEQUENCE 697 AA; 76475 MW; 58841AA4A17101AD CRC64;

Query Match 62.4%; Score 68; DB 1; Length 697;
 Best Local Similarity 66.7%; Pred. No. 1.5;
 Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAEEAAAAA 25
 DB 408 QAAQAQAQAQAQAQAQAQAQA 431

RESULT 10
 ID HR38 DROME STANDARD; PRT; 1073 AA.
 AC P49869; O18383; Q9V1K4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable nuclear hormone receptor HR38 (dHR38).

GN HR38 OR NR4A4 OR CG1864.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).
 RC TISSUE=Larva;
 RX MEDLINE=95372400; PubMed=7644522;
 RA Sutherland J.D., Kozlova T., Tzertzinis G., Kafatos F.C.;
 RT "Drosophila hormone receptor 38: a second partner for Drosophila USP
 RT suggests an unexpected role for nuclear receptors of the nerve growth
 RT factor-induced protein B type.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7966-7970(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.
 RX MEDLINE=98370123; PubMed=9704500;
 RA Komonyi O., Mink M., Caha J., Maroy P.;
 RT "Genomic organization of dHR38 gene in Drosophila: presence of
 RT Alu-like repeat in a translated exon and expression during embryonic
 RT development.";
 RL Arch. Insect Biochem. Physiol. 38:185-192(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND CHARACTERIZATION.
 RC TISSUE=Larva;
 RX MEDLINE=98315108; PubMed=9649534;
 RA Kozlova T., Pokholkova G.V., Tzertzinis G., Sutherland J.D.,
 RA Zhmulev I.F., Kafatos F.C.;
 RT "Drosophila hormone receptor 38 functions in metamorphosis: a role in
 RT adult cuticle formation.";
 RL Genetics 149:1465-1475(1998).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM LONG).
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galle R.E.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Gordon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abghyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D., Heitman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svrlaks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."


```

RL [5] Sequence:287:2185-2195 (2000).
RN
RP STRAIN=Canton-S;
RX MEDLINE=96068664; PubMed=7479849;
RA Fisk G.J., Thummel C.S.;
RT "Isolation, regulation, and DNA-binding properties of three
RL Drosophila nuclear hormone receptor superfamily members.";
CC Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608 (1995).
CC
CC -1- FUNCTION: BINDS TO NGF1-B RESPONSE ELEMENTS. PLAYS AN IMPORTANT
CC -1- SUBUNIT: FORMS A HETEROIDIMER WITH USP.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P49869-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P49869-2; Sequence=VSP_003714;
CC
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN PREBLASTODERM
CC EMBRYOS, SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL
CC TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND
CC BRAIN COMPLEXES, BUT NOT IN OVARIES.
CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.
CC HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.
CC SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN
CC LARVAE.
CC
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC subfamily.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL, X89246; CAA61534.1; -.
CC EMBL, Y15606; CAA75690.1; -.
CC EMBL, AJ002073; CAA05172.1; -.
CC EMBL, AE003667; AAF53914.1; -.
CC EMBL, U36762; AAC46926.1; -.
CC HSSB, P19793; 2NUL.
CC
CC TRANSFAC; T02760; -.
CC DR P1yBase; FB90014859; Ht38.
CC DR InterPro; IPR000536; Hormone_rec_1lg.
CC DR InterPro; IPR001723; Strdhmn_receptor.
CC DR InterPro; IPR001628; ZnF_CsteroId.
CC DR Pfam; PF00104; hormone_rec_1.
CC DR Pfam; PF00105; zf-C4; 1.
CC DR PRINTS; PR00398; STRDHOMRNER.
CC DR PRINTS; PR00047; STEROIDRNGER.
CC DR ProDom; PD000035; ZnF_CsteroId; 1.
CC
CC SMART; SM00430; HOL1_1.
CC DR SMART; SM00399; ZnF_C4; 1.
CC DR Receptor; PS00031; NOCLEAR_RECEPTOR; 1.
CC KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Alternative splicing; Developmental protein.
CC
CC FT DNA_BIND 744 809
CC FT NUCLEAR_RECEPTOR-TYPE.
CC FT ZN_FING 744 764
CC FT C4-TYPE.
CC FT ZN_FING 780 804
CC FT C4-TYPE.
CC FT ZN_FING 188 192
CC FT POLY-ALA.
CC FT DOMAIN 206 218
CC FT POLY-ALA.
CC FT DOMAIN 221 228
CC FT POLY-ALA.
CC FT DOMAIN 268 272
CC FT POLY-THR.
CC FT DOMAIN 294 312
CC FT POLY-GLN.
CC FT DOMAIN 441 462
CC FT POLY-GLN.
CC FT DOMAIN 505 508
CC FT POLY-SER.
CC FT DOMAIN 619 626
CC FT POLY-GLN.
CC FT DOMAIN 661 665
CC FT POLY-ALA.
CC FT VASAPLIC 1 522
CC Missing (in isoform Short).
CC /FTId=VSP_003714.

```

FT	CONFLICT	667	667	V -> VSSPV (IN REF. 4).
FT	CONFLICT	685	685	S -> L (IN REF. 1 AND 3).
FT	CONFLICT	689	692	STAO -> LIGER (IN REF. 2).
FT	CONFLICT	697	697	A -> D (IN REF. 2).
FT	CONFLICT	702	702	N -> S (IN REF. 2).
FT	CONFLICT	1041	1041	S -> R (IN REF. 2).
FT	CONFLICT	1064	1064	E -> D (IN REF. 2).
SO	SEQUENCE	1073 AA;	116991 MM;	126A30DAFAC096A CRC64;
Query Match				
Best Local Similarity		62.4%;	Score 68;	DB 1;
Matches 17;		Conservative	73.9%;	Pred. No. 2.1;
		0;	Mismatches	6;
			Indels	0;
			Gaps	0;
Oy	3 AAAAAAAAAAAAAAAAAAAAAA 25			
Db	206 AATAAATGAAAAAGGAASAAAA 228			
RESULT 11				
PMXB_HUMAN				
ID	PMXB_HUMAN	STANDARD;	PRT;	314 AA.
AC	Q99453;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)			
DE	(PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).			
GN	PHOX2B OR PMX2B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Neuroblastoma;			
RX	MEDLINE=97191543; PubMed=9039501;			
RA	Yokoyama M., Nishi Y., Yoshii J., Okubo K., Matsubara K.;			
RT	"Identification and cloning of neuroblastoma-specific and nerve			
RT	tissue-specific genes through compiled expression profiles.";			
RL	DNA Res. 3:311-320(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20487360; PubMed=11034547;			
RA	Adachi M., Browne D., Lewis E.J.;			
RT	"Paired-like homeodomain proteins Phox2a/Arx and Phox2b/NBPhox have			
RT	similar genetic organization and independently regulate dopamine			
RT	beta-hydroxylase gene transcription.";			
RL	DNA Cell Biol. 19:539-554(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99326521; PubMed=10395798;			
RA	Yokoyama M., Watanabe H., Nakamura M.;			
RT	"Genomic structure and functional characterization of NBPhox (PMX2B),			
RT	a homeodomain protein specific to catecholaminergic cells that is			
RL	involved in second messenger-mediated transcriptional activation.";			
RL	Genomics 59:40-50(1999).			
CC	-1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF SEVERAL MAJOR			
CC	NORADRENERGIC NEURON POPULATIONS, INCLUDING THE LOCUS COERULEUS.			
CC	TRANSCRIPTION FACTOR WHICH COULD DETERMINE A NEUROTRANSMITTER			
CC	PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED			
CC	ACTIVATION OF THE DOPAMINE BETA-HYDROXYLASE AND C-FOS PROMOTERS,			
CC	AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND			
CC	SERUM-RESPONSE ELEMENT.			
CC	-1- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL			
CC	GLAND.			
CC	-1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC				

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D82344; BA01555.1; -

DR EMBL; AF117979; AAD26698.1; -

DR EMBL; AB015671; BAA82670.1; -

DR PIR; JC5273; JC5273.

DR HSSP; P06601; 1FTL.

DR TRNSPAC; T03961; -

DR Genew; HGNC:9143; PHOX2B.

DR MIM; 603851; -

DR GO; GO:0003712; F:transcription cofactor activity; TAS.

DR GO; GO:0003700; F:transcription factor activity; TAS.

DR GO; GO:0007399; P:neurogenesis; TAS.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR007104; Paired homeo.

DR Pfam; PF00046; homeobox; 1.

DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS00071; HOMEBOX_2; 1.

DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;

KW Transcription regulation; Activator.

FT DNA BIND 98 157 HOMEBOX.

FT DOMAIN 159 167 POLY-ALA.

FT DOMAIN 212 217 POLY-GLY.

FT DOMAIN 241 260 POLY-ALA.

SO SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
Best Local Similarity 78.3%; Pred. No. 1;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
Db 244 AAAAAAAAAAAAAAAAAAGLAAA 266

RESULT 12
PMXB_MOUSE STANDARD; PRT; 314 AA.
AC 035690;
ID PMXB_MOUSE STANDARD; PRT; 314 AA.
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
OS PHOX2B OR PMX2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98040559; PubMed=9374403;
RA Pattyn A., Morin X., Cremer H., Goridis C., Brunet J.-F.;
RT "Expression and interactions of the two closely related homeobox
genes Phox2a and Phox2b during neurogenesis.";
RL Development 124:4065-4075(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99326521; PubMed=10395798;
RA Yokoyama M., Watanabe H., Nakamura M.;
RT "Genomic structure and functional characterization of NBPhox (PMX2B),
a homeodomain protein specific to catecholaminergic cells that is
involved in second messenger-mediated transcriptional activation.";
RL Genomics 59:40-50(1999).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Y14493; CAAT4833.1; -

DR EMBL; AB015672; BAA82671.1; -

DR HSSP; P06601; 1FTL.

DR TRNSPAC; T03976; -

DR MGD; MGI:1100882; Phox2b.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR007104; Paired homeo.

DR Pfam; PF00046; homeobox; 1.

DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS00071; HOMEBOX_2; 1.

DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.

FT DNA BIND 98 157 HOMEBOX.

FT DOMAIN 159 167 POLY-ALA.

FT DOMAIN 212 217 POLY-GLY.

FT DOMAIN 241 260 POLY-ALA.

SO SEQUENCE 314 AA; 31621 MW; 40737F71948B595A CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
Best Local Similarity 78.3%; Pred. No. 1;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
Db 244 AAAAAAAAAAAAAAAAAAGLAAA 266

RESULT 13
MAZ_HUMAN STANDARD; PRT; 477 AA.
ID MAZ_HUMAN STANDARD; PRT; 477 AA.
AC P56270; Q15703; Q99443;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Myc-associated zinc finger protein (MAZ) (Purine-binding
DE transcription factor) (Pur-1) (ZP87) (ZIF87).
GN MAZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92366479; PubMed=1502157;
RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
RT "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
RT regulating transcriptional initiation and termination.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92232709; PubMed=1567856;
RA Pyrc J.J., Moberg K.H., Hall D.J.;
RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds
RT to two sites within the c-myc promoter.";
RL Biochemistry 31:4102-4110(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96428591; PubMed=8831693;
RA Tsuchiuchi H., Sakatsune O., Itakura K., Yokoyama K.K.;
RT "Members of the MAZ family: a novel cDNA clone for MAZ from human
RT pancreatic islet cells.";
RL Biochem. Biophys. Res. Commun. 226:801-809(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=96224025; PubMed=8626793;

RA Parks C.L., Shenk T.;
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 RL responds to MAZ and Sp1.";
 J. Biol. Chem. 271:4417-4430(1996).
 (5)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Lymphoblastoma;
 RX MEDLINE=98352105; PubMed=9685418;
 RA Song J., Murakami H., Teutsui H., Tang X., Matsumura M., Itakura K.,
 Kazanawa I., Sun K., Yokoyama K.K.;
 "Genomic organization and expression of a human gene for Myc-
 associated zinc finger protein (MAZ).";
 J. Biol. Chem. 273:20603-20614(1998).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES
 IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,
 METAL AND METAL2, WITHIN THE C-MYC PROMOTER HAVING GREATER
 AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES
 WITHIN THE PROMOTER OF THE SP1 FAMILY OF TRANSCRIPTION FACTORS.
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
 MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.
 CC -1- SIMILARITY: Contains 6 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M94046; -; NOT ANNOTATED CDS.
 DR EMBL; M93339; -; NOT ANNOTATED CDS.
 DR EMBL; D85311; BA012728.1; ALT_INIT.
 DR EMBL; U33819; AA04121.1; ALT_INIT.
 DR EMBL; AB017335; BA033064.1; -;
 DR PIR; A42170; A42170.
 DR TRANSFAC; T00490; -;
 DR TRANSFAC; T02305; -;
 DR GeneW; HGNC:6914; MAZ.
 DR MIM; 600999; -;
 DR GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.
 DR GO; GO:0006369; P:transcription termination from Pol II promoter; TAS.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 5.
 DR ProDom; PD000003; Znf_C2H2; 1.
 DR SMART; SM00355; Znf_C2H2; 6.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
 KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
 RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 150 212 C2H2-TYPE 1.
 FT ZN_FING 279 301 C2H2-TYPE 2.
 FT ZN_FING 307 329 C2H2-TYPE 3.
 FT ZN_FING 337 360 C2H2-TYPE 4.
 FT ZN_FING 366 388 C2H2-TYPE 5.
 FT ZN_FING 392 413 C2H2-TYPE 6.
 FT DOMAIN 96 108 POLY-ALA.
 FT DOMAIN 133 139 POLY-PRO.
 FT DOMAIN 157 161 POLY-ALA.
 FT DOMAIN 245 249 POLY-GLY.
 FT DOMAIN 435 449 POLY-ALA.
 FT CONFLICT 259 259 MISSING (IN REF. 3).
 FT CONFLICT 401 401 L->M (IN REF. 2 AND 4).
 FT CONFLICT 443 443 MISSING (IN REF. 3).
 SQ SEQUENCE 477 AA; 48607 MW; C04C80F32C3C6825 CRC64;

Query Match 61.5%; Score 67; DB 1; Length 477;

Best Local Similarity 77.3%; Pred. No. 1.4; Mismatches 4; Indels 0; Gaps 0;

Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

4 AAAAAAAAAAAAAAAAAAAAAA 25

|||||:|||||:|||||:|||||

Db 90 AAAAAAAAAAAAAAAAAAAAAA 111

RESULT 14

ID HMID DROAN STANDARD; PRT; 606 AA.

AC P22544;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Homeobox protein B-H1.

GN B-H1 OR OM(1D).

OS Drosophila ananassae (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7217;

RP SEQUENCE FROM N.A.

RX MEDLINE=91122048; PubMed=1671353;

RA Tanda S., Corces V.G.;

RT "Retrotransposon-induced overexpression of a homeobox gene causes

defects in eye morphogenesis in Drosophila.";

RL EMBL J. 10:407-417(1991).

CC -1- FUNCTION: Probably involved in eye morphogenesis.

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X56682; CAA40011.1; -;
 DR PIR; S13367; S13367.
 DR HSSP; P14653; 1872.
 DR TRANSFAC; T03732; -;
 DR FlyBase; FBgn012114; Dana/B-H1.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00024; HOMEOBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEOBOX_1; 1.
 DR PROSITE; PS50071; HOMEOBOX_2; 1.
 KW DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.

FT DOMAIN 23 57 HIS/GLN-RICH (OPA-REPEAT).
 FT DOMAIN 106 124 HIS/GLN-RICH (OPA-REPEAT).
 FT DOMAIN 173 193 HIS/PRO-RICH.
 FT DNA_BIND 331 390 HOMEOBOX.
 FT DOMAIN 220 248 ALA-RICH.
 FT DOMAIN 422 434 ALA-RICH.
 FT DOMAIN 450 455 ALA-RICH.
 FT DOMAIN 503 510 ALA-RICH.
 FT DOMAIN 515 521 PRO-RICH.
 SQ SEQUENCE 606 AA; 61735 MW; AA7BB86367370FBB CRC64;

Query Match 61.5%; Score 67; DB 1; Length 606;

Best Local Similarity 78.3%; Pred. No. 1.7; Mismatches 5; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

3 AAAAAAAAAAAAAAAAAAAAAA 25

220 AAAAAAAAAARNSAAAAAAAAA 242

RESULT 15

HXAD_MOUSE STANDARD; PRT; 386 AA.

ID HXAD_MOUSE

AC Q62424;

```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A13 (Hox-1.10).
GN HoxA13 OR Hox-1.10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96259555; PubMed=8673126;
RA Mortlock D.P., Post L.C., Innis J.W.;
RT "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads
RT to arrest of digital arch formation.";
RL Nat. Genet. 13:284-289(1996).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: DEFECTS IN HoxA13 ARE THE CAUSE OF HYPODACTYLY (HD), A
CC CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH
CC STRUCTURES.
CC -!- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U59322; AAB03322.1; -.
CC DR HSSP: P14653; 1872.
CC DR TRANSPAC: T03337; -.
CC DR MGD: MGI:96173; Hoxa13.
CC DR InterPro: IPR001356; Homeobox.
CC DR Pfam: PF00046; homeobox; 1.
CC DR ProDom: PD000010; Homeobox; 1.
CC DR SMART: SM00389; HOX; 1.
CC DR PROSITE: PS00027; HOMEBOX_1; 1.
CC DR PROSITE: PS00071; HOMEBOX_2; 1.
CC KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC transcription regulation.
CC KW DOMAIN 38 51 POLY-ALA.
CC FT DNA BIND 320 379 HOMEBOX.
CC FT DOMAIN 52 57 POLY-GLY.
CC FT DOMAIN 62 66 POLY-ALA.
CC FT DOMAIN 73 84 POLY-ALA.
CC FT DOMAIN 101 104 POLY-ALA.
CC FT DOMAIN 116 133 POLY-ALA.
CC FT DOMAIN 198 205 POLY-ALA.
CC SQ SEQUENCE 386 AA; 39566 MW; 2B01DCC9B1951324 CRC64;

```

Query Match 60.6%; Score 66; DB 1; Length 386;
Best Local Similarity 70.8%; Pred. No. 1.5;
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
Db 111 EAPPSAAAAAAAAAAAAAAAAAS 134

Search completed: January 30, 2004, 00:20:45
Job time : 6.6338 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 : Search time 25.8216 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-2
Perfect score: 109
Sequence: 1 CEAATAAATAAATAAATAAATAA 25

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	77.1	1038	5	Q8MOW9
2	84	77.1	2347	5	Q8INH9
3	84	77.1	2451	5	Q8VGO5
4	82	75.2	378	11	Q8RO89
5	80	73.4	301	11	Q8BJK2
6	80	73.4	1354	11	Q8BFW8
7	79	72.5	110	11	Q91WV0
8	79	72.5	332	4	Q8IZU1
9	77	70.6	265	10	Q39598
10	76	69.7	665	11	Q90XG2
11	76	69.7	1452	4	Q9H4A0
12	76	69.7	1512	4	Q9H4A1
13	75	68.8	221	5	Q9VXD3
14	72	66.1	997	5	Q9W2J2
15	71	65.1	206	10	Q39597
16	71	65.1	1216	10	Q9SWH3

17	71	65.1	1787	10	Q9M4X9	Q9M4X9 chlamydomon
18	70	64.2	246	16	Q8G3U2	Q8G3U2 bifidobacte
19	70	64.2	323	4	Q9H782	Q9H782 homo sapien
20	70	64.2	349	4	Q43856	Q43856 homo sapien
21	70	64.2	423	4	Q75400	Q75400 homo sapien
22	70	64.2	452	4	Q75400	Q75400 homo sapien
23	70	64.2	484	11	Q923T4	Q923T4 mus musculu
24	70	64.2	512	4	Q8NAP8	Q8NAP8 homo sapien
25	70	64.2	618	11	Q8BX04	Q8BX04 mus musculu
26	70	64.2	618	11	Q8BW9	Q8BW9 mus musculu
27	70	64.2	1171	3	Q9P352	Q9P352 neurospora
28	69	63.3	218	11	Q9CS14	Q9CS14 mus musculu
29	69	63.3	512	11	Q8CB12	Q8CB12 mus musculu
30	69	63.3	1222	5	Q8T8L9	Q8T8L9 drosophila
31	69	63.3	1354	5	Q8MPN4	Q8MPN4 drosophila
32	69	63.3	1668	5	Q76330	Q76330 drosophila
33	69	63.3	1669	5	Q9V727	Q9V727 drosophila
34	68	62.4	403	16	Q9JUT2	Q9JUT2 neisseria m
35	68	62.4	543	5	Q9W4F9	Q9W4F9 drosophila
36	68	62.4	545	12	Q91TR2	Q91TR2 tupala herp
37	68	62.4	1340	16	Q91JH8	Q91JH8 streptomyce
38	68	62.4	2023	5	Q9V5Z9	Q9V5Z9 drosophila
39	68	62.4	2023	5	Q96542	Q96542 drosophila
40	67.5	61.9	324	5	Q9N6K2	Q9N6K2 drosophila
41	67.5	61.9	324	5	Q9NGI7	Q9NGI7 drosophila
42	67.5	61.9	324	5	Q9NGI8	Q9NGI8 drosophila
43	67.5	61.9	324	5	Q9N6K3	Q9N6K3 drosophila
44	67.5	61.9	324	5	Q9NGI9	Q9NGI9 drosophila
45	67	61.5	324	5	Q9NGB4	Q9NGB4 drosophila

ALIGNMENTS

RESULT 1

Q8MOW9 PRELIMINARY; PRT; 1038 AA.

ID AC Q8MOW9 01-OCT-2002 (TREMBLrel. 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE SD05989P (Fragment).

GN CG7518

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Stapleton W., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Dresnek D., Fattan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,

RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,

RA Ceiniker S.,

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

RU EMBL; AY122252; AAM52764.1; -

DR FlyBase; FBgn0038108; CG7518.

DR InterPro; IPR000104; Antifreeze 1.

DR InterPro; IPR002965; P rich extensin.

DR PRINTS; PR00308; ANTIFREEZE1.

DR PRINTS; PR01217; PRICHTEXTENSIN.

FT NON TER 1

SQ SEQUENCE 1038 AA; 109059 MW; 80C935A2C6D8A276 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 1038;
Best local similarity 91.3%; Pred. No. 1.1;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25
DB 112 AAAAAAAAAAAAAAAAAAAAAA 134

RESULT 2
Q8INH9 PRELIMINARY; PRT; 2347 AA.
ID Q8INH9;
AC Q8INH9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG7518-PB.
GN CG7518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
NCBI_taxid=7227;
OK NCBI
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN
RP
SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busan D.A.,
RA Carlson J.M., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrieria S., Frise E., Galle R.F., Gary N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragay V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Peltman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
RT "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP
SEQUENCE FROM N.A.
RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Murgall C.J., Lewis S.E.,
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP
SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP
SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AE003698; AAN14338.1; -; 23BF5FC5FCAEA64 CRC64;
SQ
SEQUENCE 2347 AA; 257013 MW; 23BF5FC5FCAEA64 CRC64;
Query Match 77.1%; Score 84; DB 5; Length 2347;
Best Local Similarity 91.3%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Cy 3 AAAAAAAAAAAAAAAAAAAAAA 25
Db 1374 AAAAAAAAAAAAAAAAAAAAAA 1396
RESULT 3
Q9VG05 PRELIMINARY; PRT; 2451 AA.
ID Q9VG05;
AC Q9VG05;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG7518 protein.
GN CG7518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
NCBI_taxid=7227;
OK NCBI
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikkelsen K.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jatali B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitzkas R., Tecor C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003698; AAF54888.2;..
 DR FlyBase; FBgn0038108; CG7518.
 DR InterPro; IPR001005; MYB_DNA_binding.
 DR PROSITE; PS00037; MYB 1; 1.
 SQ SEQUENCE 2451 AA; 266959 MW; 088A2293F27481E2 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 2451;
 Best Local Similarity 91.3%; Pred. No. 2.4;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
 DB 1374 AAAAAAAAAAAAAAAAAAAAAA 1396

RESULT 4

O8R089 PRELIMINARY; PRT; 378 AA.

AC O8R089; 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Similar to hypothetical protein FLJ11618.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027193; AAH27193.1;..
 KW Hypothetical protein.
 SQ SEQUENCE 378 AA; 39456 MW; 4C3FAF0D4AC29E69 CRC64;

Query Match 75.2%; Score 82; DB 11; Length 378;
 Best Local Similarity 87.5%; Pred. No. 0.73;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 222 EAAAAAAAAAAAAAAAAAAAAA 245

RESULT 5

O8BUK2 PRELIMINARY; PRT; 301 AA.

AC O8BUK2; 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Cell division cycle 2-like 5 (Fragment).
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Body;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK083577; BAC38957.1;..
 FT NON TER 1
 SQ SEQUENCE 301 AA; 32269 MW; A6CF891DBE25E09E CRC64;

Query Match 73.4%; Score 80; DB 11; Length 301;
 Best Local Similarity 83.3%; Pred. No. 0.93;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 74 EAAAAAAAAAAAAAAAAAAAAA 97

RESULT 6

O9EPW8 PRELIMINARY; PRT; 1354 AA.

AC O9EPW8; 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Nischcharin.
 GN NISCH.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR outbred; TISSUE=Brain;
 RX MEDLINE=20571837; PubMed=11121431;
 RA Alahari S.K., Lee J.W., Juliano R.L.,
 RT "Nischcharin, a Novel Protein That Interacts with the Integrin alphas
 RT Subunit and Inhibits Cell Migration."
 RL J. Cell Biol. 151:1141-1154(2000).
 DR EMBL; AF153344; AAG42100.1;..
 DR MGD; MGI:1928323; Nisch.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007092; LRR_SDS22.
 DR Pfam; PF00560; LRR; 5.
 DR PRINTS; PRO0019; LEURICRPT.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR PROSITE; PS50504; LRR_SDS22; 1.
 SQ SEQUENCE 1354 AA; 148060 MW; 01BD676FDC1A19247 CRC64;

Query Match 73.4%; Score 80; DB 11; Length 1354;
 Best Local Similarity 83.3%; Pred. No. 3.4;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 837 EAPAAAEAPAAAEAPAAAEAPAA 860

RESULT 7

O91WMO PRELIMINARY; PRT; 110 AA.

AC O91WMO; 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical 10.7 kDa protein.
 GN A1591529.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012681; AAH12681.1; -
DR MGD; MGI:2139008; A1591529.
DR InterPro; IPR001859; RIBosomal_P2.
DR PRINTS; PR00308; ANTI_FERZEI.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Hypothetical protein.
SQ SEQUENCE 110 AA; 10662 MW; 0581D2F635F87EAB CRC64;

Query Match
Best Local Similarity 72.5%; Score 79; DB 11; Length 110;
Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 23 DAAAAAAAAAAAAAAAAAAAAA 46

RESULT 8
Q81ZU1 PRELIMINARY; PRT; 332 AA.
AC Q81ZU1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Family with sequence similarity 9, member A.
GN FAM9A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2202142; PubMed=12113195;
RA Martinez-Garay I., Jablonka S., Sutarova M., Steuernagel P., Gal A.,
RA Kutsche K.;
RT "A New Gene Family (FAM9) of Low-Copy Repeats in Xp22.3 Expressed
RT Exclusively in Testis: Implications for Recombinations in This
RT Region.";
RL Genomics 80:259-267(2002).
DR EMBL; AF494343; AAN07162.1; -
SQ SEQUENCE 332 AA; 37339 MW; 92F22EG36038229C CRC64;

Query Match
Best Local Similarity 72.5%; Score 79; DB 4; Length 332;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 190 EAEFEAAAAAAAAAAAAAAAAA 213

RESULT 9
Q39598 PRELIMINARY; PRT; 265 AA.
AC Q39598;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gcgr-4 product (Fragment).
GN Gcgr-4.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=92119224; PubMed=1731966;
RA Wakarchuk W.W., Muller F.W., Beck C.F.;
RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
RT arrangements of directly repeated sequence motifs.";
RL Plant Mol. Biol. 18:143-146(1992).
DR EMBL; X17208; CAA35080.1; -
FT NON TER
SQ SEQUENCE 265 AA; 26216 MW; B35318B7377CF782 CRC64;

Query Match
Best Local Similarity 70.6%; Score 77; DB 10; Length 265;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25
DB 154 AAAAAAAAAAAKARVAAEAA 176

RESULT 10
Q9QXG2 PRELIMINARY; PRT; 665 AA.
AC Q9QXG2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Choroideremia protein.
GN CHM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA van den Hurk J.A., Huber I., van de Pol T.J., Cremers F.P.;
RT "Cloning and sequencing of the mouse choroideremia gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218084; AAF25478.1; -
DR MGD; MGI:892979; Cmm.
DR InterPro; IPR002005; Rab_GDI_REP.
DR Pfam; PF00996; GDI_1.
DR PRINTS; PR00891; RABGDIREF.
SQ SEQUENCE 665 AA; 73976 MW; FF71A74AD3FBDE0A CRC64;

Query Match
Best Local Similarity 69.7%; Score 76; DB 11; Length 665;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 132 EAEAEAAAEATEAAEAEEA 155

RESULT 11
Q9H4A0 PRELIMINARY; PRT; 1452 AA.
AC Q9H4A0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CDC2L5 protein kinase.
GN CDC2L5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
RA Picard A., Callebaut I., Perre E., Genevriere A.M.;
RT "A new subfamily of high molecular mass CDC2-related kinases with

```



```

RT PITAI/VRE."
RL Biochem. Biophys. Res. Commun. 279:832-837(2001).
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AJ297710; CAC10401.1; -.
DR HSSP; P24941; 1BDH.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
KW SEQUENCE 1452 AA; 158480 MW; C7ED072368B439CB CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1452;
Best Local Similarity 79.2%; Pred. No. 8.6;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 463 EAAARAAAKAAEATKAAEAAAKA 486

RESULT 12
O9H4A1 PRELIMINARY; PRT; 1512 AA.
ID O9H4A1
AC O9H4A1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CDC2L5 protein kinase.
GN CDC2L5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
RA Picard A., Callebaut I., Perre E., Genevriere A.M.;
RT "A new subfamily of high molecular mass CDC2-related kinases with
RT PITAI/VRE."
RL Biochem. Biophys. Res. Commun. 279:832-837(2001).
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AJ297709; CAC10400.1; -.
DR HSSP; P24941; 1BDH.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
KW SEQUENCE 1512 AA; 164969 MW; 283B8D553DB57650 CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1512;
Best Local Similarity 79.2%; Pred. No. 8.9;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 463 EAAARAAAKAAEATKAAEAAAKA 486

RESULT 13
O9VXD3 PRELIMINARY; PRT; 221 AA.
ID O9VXD3
AC O9VXD3;

```

```

DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG13012 protein.
GN CG13012.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport I.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.V., Wei M.-H., Ibegwam C.,
RA Jaitli B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Part V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter R., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003503; AAF48641.1; -.
DR FlyBase; FBGN030769; CG13012.
SQ SEQUENCE 221 AA; 22987 MW; A1B95919B167C5E2 CRC64;

Query Match 68.8%; Score 75; DB 5; Length 221;
Best Local Similarity 81.8%; Pred. No. 2.1;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 AAAAAAAAAAAAAAAAAAAAAA 25
DB 6 AAAAAAAAAAATAATVAAATAA 27

RESULT 14
O9W2J2 PRELIMINARY; PRT; 997 AA.
ID O9W2J2
AC O9W2J2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG18375 protein.

```

GN CG18375.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 OK
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt E.G., Nelson C.R., Milos G.L.G.,
 RA Abil J.F., Ashby A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck Y., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodeon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gioder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeagwae C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacel J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reihert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitkeas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN
 RN
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barzon J., An H., Baldwin D., Barzon J., Beeson K.Y., Busan D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodeon K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
 RA Ferreira S., Fries E., Galie R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibeagwae C., Jalili M., Kruse D., Li P., Matei B., Moshrefi A.,
 RA McIlroch T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacel J., Paragay V., Park S., Patel S., Pfeiffer B.,
 RA Phonanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitkeas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN
 RN
 RP SEQUENCE FROM N.A.
 RA Mista S., Croby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tuzy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Fries E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn J., Milburn J., Russo S.,
 RA Seattle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RA
 RP
 RP
 RA
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003453; AAF46699.2; --
 DR HSSP: Q13625; 1YCS.
 DR FlyBase: FBgn0034606; CG18375.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR001014; Antifreeze_1.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00023; ank; 2.
 DR PRINTS: PR01415; ANKYRIN.
 DR PRINTS: PR00308; ANTIPEPZET.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00248; ANK; 2.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 2.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 997 AA; 107821 MW; E712D400C2C4FD3D CRC64;

Query Match 66.1%; Score 72; DB 5; Length 997;
 Best Local Similarity 79.2%; Pred. No. 15;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEAAAAEEAAAAEAAAAA 25
 DB 432 EAAAAAEEAAAAEEAAAAEAAAAA 455

RESULT 15

ID Q39597 PRELIMINARY; PRT; 206 AA.
 AC Q39597;
 DT 01-NOV-1996 (TREMBLrel. 01. Created)
 DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23. Last annotation update)
 DE Gcgr-1 product (fragment).
 DE CGCR-1.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=cw15 mt-;
 RX MEDLINE=92119224; PubMed=1731966;
 RA Wakarchuk W.W., Muller F.W., Beck C.F.;
 RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
 arrangements of directly repeated sequence motifs.";
 RL Plant Mol. Biol. 18:143-146(1992).
 DR EMBL: X17207; CAA35079.1; --
 FT NON TER 1 1
 FT 206 206
 SQ SEQUENCE 206 AA; 19869 MW; ED3F120EF8EFAE1 CRC64;

Query Match 65.1%; Score 71; DB 10; Length 206;
 Best Local Similarity 75.0%; Pred. No. 4.7;
 Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEAAAAEEAAAAEAAAAA 25
 DB 49 EAAAAAEEAAAAEEAAAAEAAAAA 72

Fri Jan 30 06:18:16 2004

us-09-461-684C-2.rpt

Page 7

Search completed: January 30, 2004, 00:24:39
Job time : 27.8216 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 10.2535 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-3

Perfect score: 143
Sequence: 1 GGLFGAIGFIENGWEGMIDGMYG 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	93.7	550	1 HMTV52	hemagglutinin prec
2	134	93.7	550	1 HMTV53	hemagglutinin prec
3	134	93.7	550	1 HMTV77	hemagglutinin prec
4	134	93.7	550	1 HMTV80	hemagglutinin prec
5	134	93.7	550	1 HMTV33	hemagglutinin prec
6	134	93.7	550	1 HMTV89	hemagglutinin prec
7	134	93.7	550	1 HMTV21	hemagglutinin prec
8	134	93.7	550	1 HMTV98	hemagglutinin prec
9	134	93.7	550	1 HMTV15	hemagglutinin prec
10	134	93.7	550	2 JQ1153	hemagglutinin prec
11	134	93.7	550	2 JQ1154	hemagglutinin prec
12	134	93.7	550	2 JQ1155	hemagglutinin prec
13	134	93.7	566	1 HMTV8	hemagglutinin prec
14	134	93.7	566	1 HMTV8A	hemagglutinin prec
15	134	93.7	566	1 HMTV8M	hemagglutinin prec
16	134	93.7	566	1 HMTV8U	hemagglutinin prec
17	133	93.0	561	1 HMTV49	hemagglutinin prec
18	133	93.0	561	1 HMTV84	hemagglutinin prec
19	132	92.3	565	1 HMTV81	hemagglutinin prec
20	132	92.3	565	1 HMTV83	hemagglutinin prec
21	132	92.3	566	1 HMTV6	hemagglutinin prec
22	132	92.3	567	1 HMTV6	hemagglutinin prec
23	131	91.6	362	2 S38637	hemagglutinin - in
24	131	91.6	550	1 HMTV86	hemagglutinin prec
25	131	91.6	560	1 HMTV77	hemagglutinin prec
26	131	91.6	565	1 HMTV82	hemagglutinin prec
27	131	91.6	565	1 HMTV84	hemagglutinin prec
28	131	91.6	565	1 HMTV85	hemagglutinin prec
29	131	91.6	565	1 HMTV86	hemagglutinin prec

30	131	91.6	565	1 HMTV87	hemagglutinin prec
31	131	91.6	565	1 HMTV88	hemagglutinin prec
32	131	91.6	565	1 HMTV89	hemagglutinin prec
33	131	91.6	565	1 HMTV8T	hemagglutinin prec
34	131	91.6	565	1 HMTV8E	hemagglutinin prec
35	131	91.6	565	2 S33703	hemagglutinin - in
36	131	91.6	570	1 A45591	hemagglutinin prec
37	131	91.6	570	2 S22013	hemagglutinin prec
38	131	91.6	570	2 S22014	hemagglutinin prec
39	131	91.6	570	2 S22015	hemagglutinin prec
40	131	91.6	570	2 S22016	hemagglutinin prec
41	131	91.6	570	2 S22017	hemagglutinin prec
42	131	91.6	570	2 S22018	hemagglutinin prec
43	131	91.6	570	2 S22020	hemagglutinin prec
44	131	91.6	570	2 S22021	hemagglutinin prec
45	131	91.6	570	2 S22029	hemagglutinin prec

ALIGNMENTS

```
RESULT 1
HMTV52      hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C:Species:  influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: A29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: A29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19056; NID:g324208
A>Note: the sequence in Genbank entry FLAHAP, release 106, (PID:g324209) differs from tl
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein, hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:130-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-556/Product: transmembrane #status predicted <TM>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52,277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match      93.7%: Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 GGLFGAIGFIENGWEGMIDGMYG 24
Db      330 GGLFGAIGFIENGWEGMIDGMYG 352

RESULT 2
HMTV53      hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C:Species:  influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: B29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: B29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19057; NID:g324210
A>Note: the sequence in Genbank entry FLAHAPB, release 106, (PID:g324211) differs from tl
C:Genetics:
A:Map position: segment 4
```

C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F.1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F.300-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F.520-536/Domains: transmembrane #status predicted <TM1>
 F.8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted
 F.14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F.539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 3

HMI177 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: A27813
 R:Kida, H.; Kawako, Y.; Naeve, C.W.; Webster, R.G.
 Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: A27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16737; NID:G324081; PIDN:AAA3143.1; PID:G324082

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F.1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F.330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F.520-536/Domains: transmembrane #status predicted <TM1>

F.8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

F.14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F.539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 4

HMI180 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
 C:Accession: B27813

R:Kida, H.; Kawako, Y.; Naeve, C.W.; Webster, R.G.
 Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: B27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16738; NID:G324083

A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F.1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F.330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F.520-536/Domains: transmembrane #status predicted <TM1>
 F.8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted
 F.14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F.539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 5

HMI173

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C:Accession: C27813

R:Kida, H.; Kawako, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: C27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16739; NID:G324085; PIDN:AAA3145.1; PID:G324086

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F.1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F.330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F.520-536/Domains: transmembrane #status predicted <TM1>

F.8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

F.14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F.539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 6

HMI189

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C:Accession: D27813

R:Kida, H.; Kawako, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: D27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16740; NID:G324087; PIDN:AAA3146.1; PID:G324088

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F.1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F.330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F.520-536/Domains: transmembrane #status predicted <TM1>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGMYG 24
Db 330 GLFGAIAGFIENGWEGMIDGMYG 352

RESULT 7

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/62) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: E27813

R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: E27813

A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>

A:Cross-references: GB:M16741; NID:9324089
C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domains: transmembrane #status predicted <TM1>
F:7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGMYG 24
Db 330 GLFGAIAGFIENGWEGMIDGMYG 352

RESULT 8

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: F27813

R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: F27813

A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>

A:Cross-references: GB:M16742; NID:9324091
C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGMYG 24
Db 330 GLFGAIAGFIENGWEGMIDGMYG 352

RESULT 9

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: G27813

R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458; PMID:2440178

A:Accession: G27813

A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>

A:Cross-references: GB:M16743; NID:9324093; PIDN:AAA3149.1; PID:9324094
C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGMYG 24
Db 330 GLFGAIAGFIENGWEGMIDGMYG 352

RESULT 10

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01153

R:Yaounda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491; PMID:1875195

A:Accession: J01153

A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00929; NID:9221279; PIDN:BA00769.1; PID:9221280
A:Note: the authors translated the codon GGG for residue 218 as Glu

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGMYG 24

Db 330 GLFGAIAFIENGWEGMIDGWTG 352

RESULT 11

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
 C:Accession: JQ1154
 R:Yaounda, U.; Shortridge, K.F.; Shintzu, Y.; Kida, H.
 J. Gen. Virol. 72, 2007-2010, 1991
 A>Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
 A:Reference number: JQ1153; PMID:91341491; PMID:1875195
 A:Accession: JQ1154
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <YAS>
 A:Cross-references: GB:D00930; NID:9221273; PIDN:BA00770.1; PID:9221274
 A>Note: the authors translated the codon GCG for residue 218 as Glu
 A>Note: residues 528-532 are not shown in this publication
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein, homotrimer
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
 F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1,4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAFIENGWEGMIDGWTG 24

Db 330 GLFGAIAFIENGWEGMIDGWTG 352

RESULT 12

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
 C:Accession: JQ1155
 R:Yaounda, U.; Shortridge, K.F.; Shintzu, Y.; Kida, H.
 J. Gen. Virol. 72, 2007-2010, 1991
 A>Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
 A:Reference number: JQ1153; PMID:91341491; PMID:1875195
 A:Accession: JQ1155
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <YAS>
 A:Cross-references: GB:D00931; NID:9221277; PIDN:BA00771.1; PID:9221278
 A>Note: the authors translated the codon GCG for residue 218 as Glu, GCC for residue 538
 A>Note: residues 528-532 are not shown in this publication
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein, homotrimer
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
 F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1,4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAFIENGWEGMIDGWTG 24

Db 330 GLFGAIAFIENGWEGMIDGWTG 352

RESULT 13

hemagglutinin precursor - influenza A virus
 C:Species: influenza A virus
 C>Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
 C:Accession: A93705; A93233; A04051; A93231; A94441

R:Boch, G.W.; Sleight, M.J.
 Nucleic Acids Res. 8, 2561-2575, 1980

A>Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza A
 A:Reference number: A93705; PMID:81053698; PMID:6253883

A:Accession: A93705

A:Molecule type: genomic RNA

A:Residues: 1-566 <BOT>

A:Cross-references: GB:V01103

A:Experimental source: strain A/NT/60/68/29C

A>Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/60/

R:Dopheide, T.A.; Ward, C.W.

FBS Lett. 110, 181-183, 1980

A>Title: The disulfide bonds of a Hong Kong influenza virus hemagglutinin.

A:Reference number: A91276; PMID:80179105; PMID:6768586

A:Contents: annotation; disulfide bonds

R:Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.

Nature 287, 301-306, 1980

A>Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from

A:Reference number: A93233; PMID:81030852; PMID:7421990

A:Accession: A93233

A:Molecule type: genomic RNA

A:Residues: 1-24, 'S', 26, 'D', 28-159, 'G', 161-197, 'I', 199-241, 'L', 243-249 <GET>

A:Experimental source: strain X-31(H3)

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin, homotrimer; lipoprotein; thiolester bond

F:1-16/DNA: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:30-482,58-293,80-92,155-489,297-321/Disulfide bonds: #status experimental

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1,4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAFIENGWEGMIDGWTG 24

Db 346 GLFGAIAFIENGWEGMIDGWTG 368

RESULT 14

hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)

N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
 C:Accession: A93231; A04051
 R:Verhoeven, M.; Pang, R.; Min Jou, W.; Devos, R.; Huybrecock, D.; Saman, B.; Piers, W.
 Nature 286, 771-776, 1980
 A>Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A
 A:Reference number: A93231; PMID:80254693; PMID:7402351
 A:Accession: A93231
 A:Molecule type: genomic RNA
 A:Residues: 1-566 <VER>
 A:Cross-references: GB:J02090; NID:9324131; PIDN:AA43178.1; PID:9324132
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: hemagglutinin, homotrimer; lipoprotein; thiolester bond
 F:1-16/DNA: signal sequence #status predicted <SIG>
 F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
 F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
 F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1,4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAFIENGWEGMIDGWTG 24

Db 346 GLFGAIAFIENGWEGMIDGWTG 368

RESULT 15

HMLVHM
 hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000
 C:Accession: A94441; A04051
 R:Seigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
 in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, E
 A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of c
 A:Reference number: A94441
 A:Accession: A94441
 A:Molecule type: genomic RNA
 A:Residues: 1-566 <SLR>
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
 F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
 F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

 Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 GLFGAIAGFIENGWEGMIDGMYG 24
 |||||
 Db 346 GLFGAIAGFIENGWEGMIDGMYG 368
 |||||

Search completed: January 30, 2004, 00:26:21
 Job time : 10.2535 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 5.40845 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-3
Perfect score: 143
Sequence: 1 CGLFGALAGFIENGMEGMIDGMYG 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	134	93.7	550 1 HEMA_IADH1	P15582 influenza a
2	134	93.7	550 1 HEMA_IADH2	P15583 influenza a
3	134	93.7	550 1 HEMA_IADH3	P15584 influenza a
4	134	93.7	550 1 HEMA_IADH4	P15585 influenza a
5	134	93.7	550 1 HEMA_IADH5	P15586 influenza a
6	134	93.7	550 1 HEMA_IADH6	P15587 influenza a
7	134	93.7	550 1 HEMA_IADH7	P15588 influenza a
8	134	93.7	550 1 HEMA_IADH8	P43257 influenza a
9	134	93.7	550 1 HEMA_IADH9	P43258 influenza a
10	134	93.7	550 1 HEMA_IADH10	P43260 influenza a
11	134	93.7	550 1 HEMA_IADH11	P11133 influenza a
12	134	93.7	550 1 HEMA_IADH12	P11134 influenza a
13	134	93.7	556 1 HEMA_IADH13	P03437 influenza a
14	134	93.7	556 1 HEMA_IADH14	P26134 influenza a
15	134	93.7	556 1 HEMA_IADH15	P03442 influenza a
16	134	93.7	556 1 HEMA_IADH16	P26138 influenza a
17	134	93.7	556 1 HEMA_IADH17	P03449 influenza a
18	134	93.7	556 1 HEMA_IADH18	P03439 influenza a
19	134	93.7	556 1 HEMA_IADH19	P03436 influenza a
20	133	93.0	561 1 HEMA_IADH20	P15581 influenza a
21	133	93.0	561 1 HEMA_IADH21	P12439 influenza a
22	132	92.3	565 1 HEMA_IADH22	P17000 influenza a
23	132	92.3	565 1 HEMA_IADH23	P17002 influenza a
24	132	92.3	566 1 HEMA_IADH24	P03440 influenza a
25	132	92.3	566 1 HEMA_IADH25	P26139 influenza a
26	132	92.3	567 1 HEMA_IADH26	P03435 influenza a
27	131	91.6	550 1 HEMA_IADH27	P15589 influenza a
28	131	91.6	560 1 HEMA_IADH28	P03458 influenza a
29	131	91.6	565 1 HEMA_IADH29	P16994 influenza a
30	131	91.6	565 1 HEMA_IADH30	P16995 influenza a
31	131	91.6	565 1 HEMA_IADH31	P16996 influenza a
32	131	91.6	565 1 HEMA_IADH32	P16996 influenza a
33	131	91.6	565 1 HEMA_IADH33	P16558 influenza a

34	131	91.6	565 1 HEMA_IADH34	P16997 influenza a
35	131	91.6	565 1 HEMA_IADH35	P16998 influenza a
36	131	91.6	565 1 HEMA_IADH36	P16999 influenza a
37	131	91.6	565 1 HEMA_IADH37	P08011 influenza a
38	131	91.6	565 1 HEMA_IADH38	P17001 influenza a
39	131	91.6	566 1 HEMA_IADH39	P26141 influenza a
40	131	91.6	570 1 HEMA_IADH40	P26094 influenza a
41	131	91.6	570 1 HEMA_IADH41	P26095 influenza a
42	131	91.6	570 1 HEMA_IADH42	P26096 influenza a
43	131	91.6	570 1 HEMA_IADH43	P26097 influenza a
44	131	91.6	570 1 HEMA_IADH44	P26098 influenza a
45	131	91.6	570 1 HEMA_IADH45	P26099 influenza a

ALIGNMENTS

```

RESULT 1
HEMA_IADH1 STANDARD; PRT; 550 AA.
AC P15582; Q84021; Q84022;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCHI_TaxID=11357;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265456; PubMed=2440178;
RA Kida H., Kawoka Y., Nave C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119 (1987).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC -!- CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16737; AAA43143.1; -.
DR HSSP; P03437; 3HNG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 22 8 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61705 MW; 757ACFE716FC969A CRC64;
Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 2 GLFGAIGFIENGEGMIDGWTG 24
 DB 330 GLFGAIGFIENGEGMIDGWTG 352

RESULT 2

HEMA_IADH2 STANDARD; PRT; 550 AA.

AC P12583; Q84011; (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 HA. Hemagglutinin HAZ chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/8/80).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 ducks";
 RT Virology 159:109-119(1987).
 RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M16738; AAA43144.1; -
 DR PIR; B27813; HMI180.
 DR HSSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTIN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1
 FT CHAIN 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 330 550 HEMAGGLUTININ HAZ CHAIN.
 FT CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 137 K -> N (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC35 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24
 DB 330 GLFGAIGFIENGEGMIDGWTG 352

RESULT 3
 HEMA_IADH3 STANDARD; PRT; 550 AA.

AC P12584; Q84012; Q89793;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 HA. Hemagglutinin HAZ chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 ducks";
 RT Virology 159:109-119(1987).
 RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M16739; AAA43145.1; -
 DR HSSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTIN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1
 FT CHAIN 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 330 550 HEMAGGLUTININ HAZ CHAIN.
 FT CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61577 MW; 6C30BP67CFDCB7DE CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24
 DB 330 GLFGAIGFIENGEGMIDGWTG 352

RESULT 4
 HEMA_IADH4 STANDARD; PRT; 550 AA.

AC P12585; Q84013; Q84014;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 HA. Hemagglutinin HAZ chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/7/82).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11360;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoka Y., Naeye C.W., Webster R.G.;
RT Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16740; AAA43146.1; -.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutnin; Glycoprotein.
KM NON TER 1 328
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 36 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61664 MW; A16B2C8CBBD9D0 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWGMIDGWTG 24
DB 330 GLFGAIAGFIENGWGMIDGWTG 352

RESULT 5
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5 STANDARD; PRT; 550 AA.
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_Taxid=11361;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoka Y., Naeye C.W., Webster R.G.;
RT Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

```

```

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16741; AAA43147.1; -.
DR HSSP; P03437; 2VIU.
DR PIR; E27813; HMI21.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutinin; Glycoprotein.
KM NON TER 1 328
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 7 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 7 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 178 179 K -> T (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C67A15BFC CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWGMIDGWTG 24
DB 330 GLFGAIAGFIENGWGMIDGWTG 352

RESULT 6
HEMA_IADH6 STANDARD; PRT; 550 AA.
ID HEMA_IADH6 STANDARD; PRT; 550 AA.
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_Taxid=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoka Y., Naeye C.W., Webster R.G.;
RT Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M46742; AAA43148.1; -.

CC PIR: F27813; HMTV98.

DR HSP: P03437; HICJ.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR Prodom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT CHAIN 1 1 HEMAGGLUTININ HAI CHAIN.

FT NON TER 1 1

FT CAROXYD 330 550 HEMAGGLUTININ HAI CHAIN.

FT CAROXYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).

FT SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
|||
330 GLFGAIAGFIENGWEGMIDGWYG 352

DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 7

HEMA_IADH7 STANDARD; PRT; 550 AA.

ID HEMA_IADH7

AC P12588; Q84018; Q89470;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain; Hemagglutinin HAZ chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/10/85).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza A viruses; Influenzavirus A.

OX NCBI_TaxID=11363;

RN NCB1

RP SEQUENCE FROM N.A.

RX MEDLINE=87265458; PubMed=2440178;

RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild ducks".

RT Virology 159:109-119(1987).

RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HAI AND HAZ) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M16743; AAA43149.1; -.

DR HSP: P03437; 3HMG.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR Prodom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON TER 1 1

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CHAIN 330 550 HEMAGGLUTININ HAZ CHAIN.

FT CAROXYD 8 8

FT CAROXYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
|||
330 GLFGAIAGFIENGWEGMIDGWYG 352

DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 8

HEMA_IADHK STANDARD; PRT; 550 AA.

ID HEMA_IADHK

AC P43257;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain; Hemagglutinin HAZ chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hong Kong/7/75).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza A viruses; Influenzavirus A.

OX NCBI_TaxID=11364;

RN NCB1

RP SEQUENCE FROM N.A.

RX MEDLINE=91341491; PubMed=1875195;

RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;

RT "Molecular evidence for a role of domestic ducks in the introduction of avian H3 influenza viruses to pigs in southern China, where the

RT J. Gen. Virol. 72:2007-2010(1991).

RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HAI AND HAZ) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: D00929; BAA00769.1; -.

DR HSP: P03437; 2V1U.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR Prodom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CHAIN 330 550 HEMAGGLUTININ HAZ CHAIN.

FT CAROXYD 8 8

FT CAROXYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CAROXYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 550 AA; 61549 MW; 864639B829FE1BA9 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWTG 24
 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 9
 HEMA_IADHL STANDARD; PRT: 550 AA.
 AC P43258;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hong Kong/54/76).
 OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=45412;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 RT of avian H3 influenza viruses to pigs in southern China, where the
 RT A/Hong Kong/68 (H3N2) strain emerged.";
 RL J. Gen. Virol. 72:2007-2010(1991)
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D00931; BAA00771.1; -.
 DR HSSP: P03437; 2VUJ.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein.
 KW NON TER
 FT CHAIN 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61718 MW; A351C56789E4B9A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWTG 24
 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 10
 HEMA_IAGHK STANDARD; PRT: 550 AA.
 AG P43260;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
 OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=45414;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 RT of avian H3 influenza viruses to pigs in southern China, where the
 RT A/Hong Kong/68 (H3N2) strain emerged.";
 RL J. Gen. Virol. 72:2007-2010(1991).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D00930; BAA00770.1; -.
 DR HSSP: P03437; 2VUJ.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein.
 KW NON TER
 FT CHAIN 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWTG 24
 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 11
 HEMA_IAGHZ STANDARD; PRT: 550 AA.
 ID HEMA_IAGHZ
 AC P11133; Q84019; Q84020;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
 DE chain] (Fragment).
 GN HA.

OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
 CC Cell RECEPTORS AND FOR INITIATING INFECTION.
 CC Influenza A viruses; Orthomyxoviridae;
 CC NCBI_TaxId=11497;
 CC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 in China."
 RT Virology 162:160-166(1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M19057; AAA43212.1; -.
 CC HSSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Hemagglutinin; Envelope protein; Glycoprotein.
 FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 1 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C53ICE8 CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLEFGAIAGFIENGWEGMIDGWYG 24
 DB 330 GLEFGAIAGFIENGWEGMIDGWYG 352
 RESULT 12
 ID HEMA_IJAZH3 STANDARD; PRT; 550 AA.
 AC P1134; Q84025; Q84026;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzaviridae;
 CC NCBI_TaxId=11498;
 CC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 in China."
 RT Virology 162:160-166(1988).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M19056; AAA43211.1; ALT_TERM.
 CC HSSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Hemagglutinin; Envelope protein; Glycoprotein.
 FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 1 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61580 MW; 991F6D8BC02F24F2 CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLEFGAIAGFIENGWEGMIDGWYG 24
 DB 330 GLEFGAIAGFIENGWEGMIDGWYG 352
 RESULT 13
 ID HEMA_IJAIC STANDARD; PRT; 566 AA.
 AC P03437;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Aichi/2/68).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CC NCBI_TaxId=150147;
 CC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80254693; PubMed=7402351;
 RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
 RA Saman E., Fiers W.;
 RT "Antigenic drift between the haemagglutinin of the Hong Kong
 RT influenza strains A/Aichi/2/68 and A/Victoria/3/75."
 RT Nature 286:771-776(1980).
 GN Nature 286:771-776(1980).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=81123029; PubMed=7464906;
 RA Wilson I.A., Skehel J.J., Wiley D.C.;
 RT "Structure of the haemagglutinin membrane glycoprotein of influenza
 RT virus at 3-A resolution."
 RT Nature 289:366-373(1981).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=88232903; PubMed=3374584;

RA Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA Wiley D.C.;
RT "Structure of the influenza virus haemagglutinin complexed with its
RT receptor, sialic acid";
RL Nature 333:426-431(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX MEDLINE=90107940; PubMed=2295311;
RA Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA Wiley D.C.;
RT "The structure of a membrane fusion mutant of the influenza virus
RT haemagglutinin";
RL EMBO J. 9:17-24(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=90230310; PubMed=2239580;
RA Weis W.I., Bruegger A.T., Skehel J.J., Wiley D.C.;
RT "Refinement of the influenza virus hemagglutinin by simulated
RT annealing";
RL J. Mol. Biol. 212:737-761(1990).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=94352388; PubMed=8072525;
RA Bullock P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT "Structure of influenza haemagglutinin at the pH of membrane fusion";
RL Nature 371:37-43(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
RX MEDLINE=96120975; PubMed=9461077;
RA Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT "Antigen distortion allows influenza virus to escape neutralization";
RL Nat. Struct. Biol. 5:119-123(1998).
CC -I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -I- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -I- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02090; AAA43178.1; -;
DR EMBL; V01085; CAA24269.1; -;
DR PIR; A93231; HMTVHA.
DR PDB; 2HMG; 31-OCT-93.
DR PDB; 3HMG; 31-OCT-93.
DR PDB; 4HMG; 31-OCT-93.
DR PDB; 5HMG; 31-JAN-94.
DR PDB; 1HGD; 31-JAN-94.
DR PDB; 1HGE; 31-JAN-94.
DR PDB; 1HGF; 31-JAN-94.
DR PDB; 1HGG; 31-JAN-94.
DR PDB; 1HGH; 31-JAN-94.
DR PDB; 1HGI; 31-JAN-94.
DR PDB; 1HGU; 31-JAN-94.
DR PDB; 1HJM; 14-FEB-95.
DR PDB; 2VIR; 29-APR-98.
DR PDB; 2VIS; 29-APR-98.
DR PDB; 2VIT; 29-APR-98.
DR PDB; 2VIU; 29-APR-98.
DR PDB; 1E08; 29-NOV-00.
DR PDB; 1HA0; 22-DEC-99.
DR PDB; 1JRH; 13-MAR-02.
DR PDB; 1KEN; 24-APR-02.
DR PDB; 1OFU; 29-DEC-99.
DR PDB; 1OUI; 05-JAN-00.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CHAIN 346 566
FT DISULFID 30 482
FT DISULFID 68 293
FT DISULFID 80 92
FT DISULFID 113 155
FT DISULFID 297 321
FT DISULFID 489 493
FT CARBOHYD 24 24
FT CARBOHYD 38 38
FT CARBOHYD 54 54
FT CARBOHYD 97 97
FT CARBOHYD 181 181
FT CARBOHYD 301 301
FT CARBOHYD 499 499
FT STRAND 27 35
FT STRAND 40 42
FT STRAND 50 52
FT STRAND 55 57
FT STRAND 59 60
FT STRAND 67 70
FT STRAND 74 76
FT TURN 78 79
FT HELIX 82 87
FT TURN 88 88
FT HELIX 90 95
FT TURN 96 97
FT STRAND 99 99
FT STRAND 102 105
FT TURN 107 108
FT STRAND 116 117
FT TURN 119 120
FT HELIX 121 131
FT STRAND 133 133
FT STRAND 136 138
FT TURN 144 145
FT STRAND 146 147
FT STRAND 152 157
FT TURN 158 159
FT STRAND 160 162
FT TURN 165 166
FT STRAND 167 169
FT TURN 171 173
FT STRAND 174 175
FT STRAND 176 176
FT STRAND 180 185
FT STRAND 192 200
FT HELIX 204 211
FT STRAND 217 221
FT STRAND 226 229
FT STRAND 239 239
FT TURN 240 241
FT STRAND 242 242
FT STRAND 245 253
FT TURN 255 256
FT STRAND 258 265
FT STRAND 267 270
FT STRAND 272 275
FT STRAND 282 285
FT STRAND 290 294
FT STRAND 297 299
FT TURN 300 301
FT STRAND 302 304
FT STRAND 310 311
FT STRAND 318 320
FT STRAND 323 324
FT STRAND 331 333
FT STRAND 337 337
FT TURN 347 348
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
N-LINKED (GLCNAC. . .)
HEMAGGLUTININ HA1 CHAIN.
HEMAGGLUTININ HA2 CHAIN.
INTERCHAIN.

FT TURN 350 350
 FT STRAND 351 351
 FT TURN 352 354
 FT STRAND 355 355
 FT STRAND 359 360
 FT TURN 361 361
 FT STRAND 367 373
 FT TURN 374 375
 FT STRAND 376 382
 FT TURN 383 400
 FT TURN 401 401
 FT STRAND 406 407
 FT HELIX 421 471
 FT STRAND 472 474
 FT STRAND 475 477
 FT STRAND 482 485
 FT HELIX 491 498
 FT TURN 499 500
 FT HELIX 504 515
 FT TURN 518 519
 SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 14
 HEMA_IADU3 STANDARD; PRT; 566 AA.

AC P26134;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Duck/Alberta/78/76).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OK NCBI_TaxID=11348;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92114135; PubMed=1731092;
 RA Bean W.J., Scheil M., Katz J., Kawoka Y., Naeve C., Gorman O.,
 RA Webster R.G.;
 RT "Evolution of the H3 influenza virus hemagglutinin from human and
 RT nonhuman hosts.";
 RL J. Virol. 66:1129-1138(1992).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M73771; -; NOT_ANNOTATED_CDS.
 DR HSPF; P03437; 2VIT.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTIN12.
 DR ProDom; PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 344
 FT CHAIN 346 566
 FT CHAIN 23 23
 FT CARBOHYD 24 24
 FT CARBOHYD 38 38
 FT CARBOHYD 54 54
 FT CARBOHYD 181 181
 FT CARBOHYD 301 301
 FT CARBOHYD 499 499
 SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FP9415B89 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 15
 HEMA_IADU3 STANDARD; PRT; 566 AA.

AC P03442;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Duck/Ukraine/1/63).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OK NCBI_TaxID=11374;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82025542; PubMed=6169439;
 RA Fang R., Min Jou W., Huylebroeck D., Devos R., Fiers W.;
 RT "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
 RT gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
 RT hemagglutinin.";
 RL Cell 25:315-323(1981).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01087; CAA24271.1; -.
 DR PDB; 1IBN; 08-AUG-01.
 DR PDB; 1IBO; 08-AUG-01.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTIN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 16
 FT CHAIN 17 344
 FT CHAIN 346 566
 FT CARBOHYD 24 24
 FT CARBOHYD 38 38
 FT CARBOHYD 54 54
 FT CARBOHYD 97 97
 HEMAGGLUTININ HA1 CHAIN.
 HEMAGGLUTININ HA2 CHAIN.
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
 |||||
 Db 346 GLFGAIGFIENGWEGMIDGWTG 368

Search completed: January 30, 2004, 00:20:45
 Job time : 5.40845 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 24.7887 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-3
Perfect score: 143
Sequence: 1 GLEFGAIGFIENGWEGMIDGWYG 24

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	134	93.7	384	12	Q8JK63 influenza a
2	134	93.7	566	12	Q98052 influenza a
3	134	93.7	566	12	Q8U251 influenza a
4	134	93.7	566	12	Q8OLN8 influenza a
5	134	93.7	566	12	Q67132 influenza a
6	134	93.7	566	12	Q67125 influenza a
7	134	93.7	566	12	Q8UKX3 influenza a
8	134	93.7	566	12	Q91MX7 influenza a
9	134	93.7	566	12	Q9DHG0 influenza a
10	134	93.7	566	12	Q910M5 influenza a
11	134	93.7	566	12	Q67126 influenza a
12	133	93.0	301	12	Q9DXE3 influenza a
13	132	92.3	550	12	Q82499 influenza a
14	132	92.3	550	12	Q82498 influenza a
15	132	92.3	550	12	Q82753 influenza a
16	132	92.3	566	12	Q82496 influenza a

17	132	92.3	571	12	Q03909 influenza a
18	131	91.6	109	12	Q67050 influenza a
19	131	91.6	109	12	Q67053 influenza a
20	131	91.6	109	12	Q67051 influenza a
21	131	91.6	109	12	Q67052 influenza a
22	131	91.6	362	12	Q9QKD3 influenza a
23	131	91.6	362	12	Q9QKD1 influenza a
24	131	91.6	362	12	Q82513 influenza a
25	131	91.6	362	12	Q9QKD2 influenza a
26	131	91.6	362	12	Q84174 influenza a
27	131	91.6	362	12	Q82517 influenza a
28	131	91.6	365	12	Q9DL25 influenza a
29	131	91.6	367	12	Q9DL22 influenza a
30	131	91.6	368	12	Q9DL29 influenza a
31	131	91.6	369	12	Q9DL26 influenza a
32	131	91.6	369	12	P87689 influenza a
33	131	91.6	369	12	Q9DL06 influenza a
34	131	91.6	371	12	Q9DL24 influenza a
35	131	91.6	371	12	P87685 influenza a
36	131	91.6	372	12	Q9DL20 influenza a
37	131	91.6	374	12	Q9DL21 influenza a
38	131	91.6	375	12	Q9DL27 influenza a
39	131	91.6	375	12	Q9DL05 influenza a
40	131	91.6	376	12	Q9DL30 influenza a
41	131	91.6	376	12	Q9DL04 influenza a
42	131	91.6	377	12	Q9E7P6 influenza a
43	131	91.6	382	12	Q9DL03 influenza a
44	131	91.6	408	12	Q9E7P5 influenza a
45	131	91.6	409	12	Q9Q0L5 influenza a

ALIGNMENTS

RESULT 1

Q8JK63 PRELIMINARY; PRT; 384 AA.

AC Q8JK63:

DT 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Hemagglutinin (Fragment).

GN H3HA.

OS Influenza A virus (A/real/Germany/wv201r/01).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

OX NCBI_TaxID=205472;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A/real/Germany/wv01r/01;

RA Werner O., Starick B., Mueller T., Muehle R.;

RT "Characterisation of avian influenza virus isolates from wild birds from Germany."

RT Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL; AJ506781; CAD44999.1;--

DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTN12.

DR ProDom; PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

FT NON TER 384

FT SEQUENCE 384 AA; 42076 MW; 459731795CA5CB38 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 384;

Best Local Similarity 100.0%; Pred. No. 5.9e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLEFGAIGFIENGWEGMIDGWYG 24

Db 346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 2

```

098052 ID 098052 PRELIMINARY; PRT; 566 AA.
AC 098052;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hemagglutinin precursor (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OK NCBI_TaxID=197911;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=81053698; PubMed=6253883;
RA Both G.W., Sleight M.J.;
RT "Complete nucleotide sequence of the haemagglutinin gene from a human
RT Influenza virus of the Hong Kong subtype.";
RL Nucleic Acids Res. 8:2561-2575(1980).
RN (2)
RP SEQUENCE OF 17-344 FROM N.A.
RX MEDLINE=81194918; PubMed=6164798;
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT subtype: Correlation of amino acid changes with alterations in viral
RT antigenicity.";
RL J. Virol. 37:845-853(1981).
RN (3)
RP SEQUENCE OF 17-566 FROM N.A.
RX MEDLINE=82033276; PubMed=6169843;
RA Both G.W., Sleight M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype
RT Influenza viruses during antigenic drift.";
RL J. Virol. 39:845-853(1981).
RN (4)
RP FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
RN CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: J02135; AAA43189.1; -.
DR HSP; P03437; HKGE.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CHAIN 346 566
FT CHAIN 566 566
SQ SEQUENCE 566 AA; 63414 MW; C447FD465BE4FCF9 CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 8.8e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 3
 ID 08U251 PRELIMINARY; PRT; 566 AA.
 AC 08U251;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE Haemagglutinin.
 GN HA.

OS Influenza A virus (A/pet bird/Hong Kong/1559/99 (H3N8)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OK NCBI_TaxID=183665;

```

RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A/pet bird/Hong Kong/1559/99;
RA Chin P., Shortridge K.F.;
RT "Characterisation of avian H3 influenza viruses.";
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ427304; CAD20336.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63403 MW; F11C91B6A0183484 CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 8.8e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 346 GLFGAIAGFIENGWEGMIDGWYG 368

```

RESULT 4
08QLN8 ID 08QLN8 PRELIMINARY; PRT; 566 AA.
AC 08QLN8;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Haemagglutinin.
GN HA.
OS Influenza A virus (A/aquatic bird/Hong Kong/399/99 (H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OK NCBI_TaxID=183664;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A/aquatic bird/Hong Kong/399/99;
RA Chin P., Shortridge K.F.;
RT "Characterisation of influenza viruses from wild aquatic birds.";
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ427297; CAD20332.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63412 MW; 68913C222C97B92E CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 8.8e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 5

```

067132
ID 067132 PRELIMINARY; PRT: 566 AA.
AC 067132;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Aichi/2/68;
RA Min J.W., Verhoeven M., Fang R.-X., Devos R., Huybrecock D.,
RA Fiers W.;
RT "Shift and drift in influenza viruses.";
RL (in) Carlile M.J., Collins J.F., Moseley B.E. B. (eds.);
RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,
RL Cambridge University Press, New York (1981)
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; M55059; AAA43239.1; -.
DR HSSP; P03437; IHGS.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT CHAIN 1 344
FT NEURAMINIDASE.
FT SEQUENCE 566 AA; 63441 MW; E5D1B97DF6FECA CRC64;
SQ
Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 346 GLFGAIGFIENGWEGMIDGWYG 368
RESULT 6
067125 PRELIMINARY; PRT: 566 AA.
AC 067125;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Seal/M/3911/92;
RA MEDLINE=95146951; PubMed=7844533;
RA Callan R.U., Early G., Kida H., Hinshaw V.S.;
RT "The appearance of H3 influenza viruses in seals.";
RL J. Gen. Virol. 76:199-203(1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; U31949; AAA64229.1; -.
DR HSSP; P03437; 2VU.

```

```

DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63456 MW; AE556302A9EBB99F CRC64;
Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 346 GLFGAIGFIENGWEGMIDGWYG 368
RESULT 7
08UXR3 PRELIMINARY; PRT: 566 AA.
AC 08UXR3;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (A/swine/Potsdam/35/82 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=183769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/swine/Potsdam/35/82;
RA Groetzinger I., Sues J., Groetzinger C.;
RT "Evolution of european human and porcine influenza viruses.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ252132; CAC81018.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63529 MW; 6AA44C84B4DDE68A CRC64;
Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 346 GLFGAIGFIENGWEGMIDGWYG 368
RESULT 8
091MA7 PRELIMINARY; PRT: 566 AA.
AC 091MA7;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (A/Hong Kong/1/68 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=108859;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Hong Kong/1/68 (H3N2);
RA MEDLINE=21287244; PubMed=11371620;

```

QY	DB	Q910MS	Q910MS	PRELIMINARY;	PRT;	566 AA.
Query Match	Best Local Similarity	93.7%;	Score 134;	DB 12;	Length 566;	
Matches	23;	Conservative	0;	Mismatches	0;	Gaps 0
2	GLFGAIAGFIENGWEGMIDGWYG	24				
346	GLFGAIAGFIENGWEGMIDGWYG	368				
RESULT 9						
Q9DHG0	Q9DHG0	PRELIMINARY;	PRT;	566 AA.		
AC	Q9DHG0;					
DT	01-MAR-2001 (TREMBLrel. 16, Created)					
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)					
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)					
DE	Haemagglutinin precursor					
OS	Influenza A virus H3N2.					
OC	Virusess: ASRNA negative-strand virusess; Orthomyxoviridae;					
OC	Influenza A virusess; Influenzavirus A.					
OX	NCBI_Taxid=41857;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=clone 7a;					
RA	Mohsin M.A., Morris S.J., Smith H., Sweet C.;					
RT	"Influenza virus-induced apoptosis: a dual role for viral					
RT	neuraminidase.";					
RL	Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.					
CC	-1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO					
CC	CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).					
CC	-1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS					
CC	(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).					
CC	-1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.					
DR	EMBL; AJ289703; CAC18525.1; -.					
DR	HSSP; P03437; 2VIU.					
DR	InterPro: IPR001364; Hemagglutn.					
DR	Pfam; PF00509; Hemagglutinin; 1.					
DR	PRINTS; PR00329; HEMAGGLUTN12.					
DR	ProDom; PD000225; Hemagglutn; 1.					
KW	Envelope protein; Glycoprotein; Hemagglutinin; Signal.					
FT	SIGNAL 1 16 POTENTIAL.					
SO	SEQUENCE 566 AA; 63356 MW; 0BA681929300F72F CRC64;					
Query Match	Best Local Similarity	93.7%;	Score 134;	DB 12;	Length 566;	
Matches	23;	Conservative	0;	Mismatches	0;	Gaps 0
2	GLFGAIAGFIENGWEGMIDGWYG	24				
346	GLFGAIAGFIENGWEGMIDGWYG	368				
RESULT 10						
Q910MS	Q910MS	PRELIMINARY;	PRT;	566 AA.		

```

AC      0910M5 ; 01-DEC-2001 (TReMBLrel. 19, Created)
DT      01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TReMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      Hemagglutinin.
OS      Influenza A virus (A/Hong Kong/1/68(H3N2)).
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza A viruses; Influenzavirus A.
OX      NCBI_TaxID=108859;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A/Hong Kong/1/68(H3N2);
RX      MEDLINE=21287244; PubMed=11371620;
RA      Brown E.G., Liu H., Kit L.C., Baird S., Negarallah M.;
RT      "Pattern of mutation in the genome of influenza A virus on adaptation
RT      to increased virulence in the mouse lung: Identification of functional
RT      themes.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC      -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC      CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC      -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC      (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR      EMBL; AF348179; AAK51721.1; -.
DR      EMBL; AF348177; AAK51720.1; -.
DR      EMBL; AF348178; AAK51720.1; -.
DR      InterPro; IPR001364; Hemagglutn.
DR      Pfam; PF00509; Hemagglutinin; 1.
DR      PRINTS; PR00329; HEMAGGLUTN12.
DR      ProDom; PD000225; Hemagglutn; 1.
KW      Envelope protein; Glycoprotein; Hemagglutinin.
SQ      SEQUENCE 566 AA; 63530 MW; 7CB9P5BAF1BE8P94 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8,8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 GLFGAIGFTENGWEGMIDGMYG 24
DB      346 GLFGAIGFTENGWEGMIDGMYG 368

RESULT 11
ID      067126 PRELIMINARY; PRT; 566 AA.
AC      067126;
DT      01-NOV-1996 (TReMBLrel. 01, Created)
DT      01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT      01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE      Hemagglutinin.
DE      HA.
GN      Influenzavirus A.
OS      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza A viruses.
OC      NCBI_TaxID=197911;
OX      [1]
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A/Seal/MA/3984/92;
RX      MEDLINE=95146951; PubMed=7844533;
RA      Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT      "The appearance of H3 influenza viruses in seals.";
RL      J. Gen. Virol. 76:199-203(1995).
CC      -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC      CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC      -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC      (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR      EMBL; U32024; AAB64228.1; -.
DR      HSSP; P03437; 2VIU.
DR      InterPro; IPR001364; Hemagglutn.
DR      Pfam; PF00509; Hemagglutinin; 1.
DR      PRINTS; PR00329; HEMAGGLUTN12.
DR      ProDom; PD000225; Hemagglutn; 1.

```


KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CEE7D08 CRC64;
 Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 8.8e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 |||||
 DB 346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 12
 O9DXE3 PRELIMINARY; PRT; 301 AA.
 AC O9DXE3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin (Fragment).
 GN HAI.
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=140665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Shorebird/Taiwan/31-4/99;
 RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shien H.K.;
 RT "Identification and subtyping of avian influenza virus by reverse
 RT transcription-polymerase chain reaction.";
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RU -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF311750; AAC3016.1. -
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON TER 1
 FT NON TER 301
 SQ SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;

Query Match. 93.0%; Score 133; DB 12; Length 301;
 Best Local Similarity 95.7%; Pred. No. 6.2e-10;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 |||||
 DB 250 GLFGAIAGFIENGWEGMIDGWYG 272

RESULT 13
 O82498 PRELIMINARY; PRT; 550 AA.
 AC O82498;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin HAI and HA2 (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82/BS;
 RA Hartley C.A., Ward A.C., Anders E.M.;
 RT "Virulence of influenza virus for mice is associated with loss of

RT oligosaccharide from the hemagglutinin molecule.";
 RU Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; U08859; AAA18782.1. -
 DR HSSP; P03437; 2YIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON TER 1
 FT CHAIN 1 328 HAI.
 FT CHAIN 330 550 HA2.
 SQ SEQUENCE 550 AA; 61772 MW; 50BD62B6BEF11FD8 CRC64;

Query Match. 92.3%; Score 132; DB 12; Length 550;
 Best Local Similarity 95.7%; Pred. No. 1.6e-09;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 |||||
 DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 14
 O82498 PRELIMINARY; PRT; 550 AA.
 AC O82498;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin HAI and HA2 (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82;
 RA Hartley C.A., Ward A.C., Anders E.M.;
 RT "Virulence of influenza virus for mice is associated with loss of
 RT oligosaccharide from the hemagglutinin molecule.";
 RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RU -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; U08858; AAA18781.1. -
 DR HSSP; P03437; 2YIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON TER 1
 FT CHAIN 1 328 HAI.
 FT CHAIN 330 550 HA2.
 SQ SEQUENCE 550 AA; 61802 MW; 11443B1CE5A1F6A CRC64;

Query Match. 92.3%; Score 132; DB 12; Length 550;

Best Local Similarity 95.7%; Pred. No. 1.6e-09;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAAGIENGWEGMIDGWYG 24
|:|||||
Db 330 GFGAAGIENGWEGMIDGWYG 352

RESULT 15

Q82753 PRELIMINARY; PRT; 550 AA.
ID Q82753;
AC Q82753;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 02, Last annotation update)
DE Haemagglutinin (Fragment).
OS Influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC unclassified Orthomyxoviridae.
OX NCBI_TaxID=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82/BS/ML10;
RX MEDLINE=97300854; PubMed=9155874;
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)
virus associated with increased virulence for mice.";
RL Arch. Virol. 142:75-88 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82/BS/ML10;
RX MEDLINE=97456249; PubMed=9311563;
RA Ward A.C.;
RT "Virulence of influenza A virus for mouse lung.";
RL Virus Genes 14:187-194 (1997).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR HMBP; U08905; AAC79579.1; -.
DR HSBP; P03437; 2VIU.
DR InterPro; IPR01364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 1
FT CHAIN 1 328 HAEMAGGLUTININ HA1.
FT CHAIN 330 550 HAEMAGGLUTININ HA2.
SQ SEQUENCE 550 AA; 61745 MW; 692A49DB678AC4BC CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 1.6e-09;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAAGIENGWEGMIDGWYG 24
|:|||||
Db 330 GFGAAGIENGWEGMIDGWYG 352

Search completed: January 30, 2004, 00:24:41
Job time : 26.7887 secs

A:Map position: 3
 A:Introns: 312/3; 359/3; 444/3
 C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match 48.8%; Score 102; DB 2; Length 517;
 Best Local Similarity 80.8%; Pred. No. 0.17;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 20 EAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 DB 440 ERVGGKKKKKKKKKKKKKKKKKKKKKK 465

RESULT 3

T39683
 znotin-like protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
 C:Accession: T39683; T40195
 R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, March 1998
 A:Reference number: 221869
 A:Accession: T39683
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-124 <OLI>

A:Cross-references: EMBL:AL049489; PIDN:CA839796.1; GSPDB:GN00067; SPDB:SPBC1778.01c
 A:Experimental source: strain 972h; cosmid c1778

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hiltbert, H.; Duesterhoeft, A.
 submitted to the EMBL Data Library, February 1998
 A:Reference number: 221910
 A:Accession: T40195

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 89-442 <WOO>
 A:Cross-references: EMBL:Z97992; PIDN:CA810796.1; GSPDB:GN00067; SPDB:SPBC30D10.01
 A:Experimental source: strain 972h; cosmid c30D10

C:Genetics:
 A:Gene: SPDB:SPBC1778.01c; SPDB:SPBC30D10.01
 A:Map position: 2

Query Match 48.6%; Score 101.5; DB 2; Length 442;
 Best Local Similarity 56.0%; Pred. No. 0.17;
 Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;

OY 2 EAAAAA-----EAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 DB 309 EAAAAAQQKKEEERRRAEAAAKASAAAANKKAKKQKRDKTVK 358

RESULT 4

152523
 nucleoporin p62 homolog - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)
 C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
 C:Accession: I52523
 R:Wang, Z.Q.; Akmal, K.M.; Kim, K.H.
 Biol. Reprod. 51, 1022-1030, 1994

A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the ger
 A:Reference number: I52523; MUID:95151924; PMID:7849178
 A:Accession: I52523

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

A:Residues: 1-215 <RES>
 A:Cross-references: GB:S75997; NID:9913245; PIDN:AAB33384.1; PID:9913246
 A:Experimental source: testis

Query Match 47.8%; Score 100; DB 2; Length 215;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKKKKKKKK 45
 |||||

DB 35 KKKKKKKKKKKKKKKKKKKKK 54

RESULT 5

T46395
 hypothetical protein DKFZp434i1120.1 - human (fragment)

C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T46395
 R:Ottenwajlder, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, January 2000
 A:Reference number: 223031
 A:Accession: T46395

A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-380 <AAA>
 A:Cross-references: EMBL:AL137556
 A:Experimental source: adult testis; clone DKFZp434i1120
 C:Genetics:
 A:Note: DKFZp434i1120.1

Query Match 47.8%; Score 100; DB 2; Length 380;
 Best Local Similarity 100.0%; Pred. No. 0.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKKKKKKKKKKKKKKKKKKKKKK 45
 DB 355 KKKKKKKKKKKKKKKKKKKKKKK 374

RESULT 6

S09388
 histone H1 - sea urchin (Parechinus angulosus)

C:Species: Parechinus angulosus (angulate urchin)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
 C:Accession: S09388
 R:Hill, C.S.; Martin, S.R.; Thomas, J.O.
 EMBO J. 8, 2591-2599, 1989

A:Title: A stable alpha-helical element in the carboxy-terminal domain of free and chromo
 A:Reference number: S09388; MUID:90060019; PMID:2583125

A:Accession: S09388
 A:Status: preliminary

A:Molecule type: protein
 A:Residues: 1-206 <HIL>
 C:Superfamily: histone H1
 C:Keywords: chromosomal protein

Query Match 46.9%; Score 98; DB 2; Length 206;
 Best Local Similarity 55.8%; Pred. No. 0.18;
 Matches 24; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAEAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 DB 140 KAAAKRKAAAKKRAAAAKRAATKAKKAKKPKKTKAKKAKK 182

RESULT 7

S25194
 znotin - yeast (Saccharomyces cerevisiae)

N:Alternate names: probable 2-DNA-binding protein; protein G9554; protein YGR285C
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
 C:Accession: S25194; S64620; S19066
 R:Zheng, S.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A.
 EMBO J. 11, 3787-3796, 1992

A:Title: Znotin, a putative 2-DNA binding protein in Saccharomyces cerevisiae.
 A:Reference number: S25194; MUID:93010971; PMID:1396572

A:Accession: S25194
 A:Molecule type: DNA

A:Residues: 1-433 <ZNA>
 A:Cross-references: EMBL:X63612; NID:94836; PIDN:CA445156.1; PID:94837
 A:Note: part of this sequence, including the amino end of the mature protein, was confirm
 R.Voet, M.; Voelckaert, G.


```
A:Accession: F71619
R:Gardner, M.J.; Telcelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
P;Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
C:Date: 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; PMID:99021743; PMID:9804551
A:Accession: F71619
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-483 <GAR>
A:Cross-references: GB:AEO01382; GB:AEO01362; NID:g3845130; PIDN:AC71836.1; PID:g384513
C:Genetics:
A:Experimental source: clone 3D7
A:Gene: PFB0235W

Query Match          43.1%; Score 90; DB 2; Length 483;
Best Local Similarity 85.0%; Pred. No. 1.4;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy      26 KKKKKKKKKKKKKKKKKK 45
       |||:|||||:::|||
Db      449 KKKKKKKKKKKKKKKKKK 468

RESULT 13
T50609
hypothetical protein DKFZp761B2423.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C/Accession: T50609
R:Blocher, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A:Reference number: Z25143
A:Accession: T50609
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-529 <AAA>
A:Cross-references: EMBL:ALJ59564
A:Experimental source: adult amygdala; clone DKFZp761B2423
C/Genetics:
A>Note: DKFZp761B2423.1

Query Match          43.1%; Score 90; DB 2; Length 529;
Best Local Similarity 45.0%; Pred. No. 1.5;
Matches 18; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Oy      5 AAABAAAAEAFAAAAKKKKKKKKKKKKKKKKKKK 44
       : : : : : : : : : : : : | | | | | |
Db     444 SGSGTRSHSSASNAESQDSKKKKKKKHKKHKKHKHX 483

RESULT 14
T23778
histone H1.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C/Accession: T23778; T42231; S09130; S01817
R:Percy, C.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19798
A:Accession: T23778
A>Status: translated from GB/EMBL/DDBT
A:Molecule type: DNA
A:Residues: 1-208 <MII>
A:Cross-references: GB:A79603; PIDN:CAB01892.1; GSFPDB:GN00028; CESP:ML63.3
A:Experimental source: clone M163
R:Jeurnusik, M.; Schulze, B.
submitted to the EMBL Data Library, August 1997
A>Description: The histone H1 complement of Caenorhabditis elegans.
A:Reference number: Z22091
A:Accession: T42231
A>Status: translated from GB/EMBL/DDBT
A:Molecule type: mRNA
```

[illegible]

```

RESULT 15
S19113
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
C|Species: Chlamydomonas reinhardtii
C|Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C|Accession: S19113; SI4466
R|Makarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A|Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A|Reference number: S19113; MUID:92119224; PMID:1731966
A|Accession: S19113
A|Status: Preliminary
A|Molecule type: DNA
A|Residues: 1-265 >(WAK>
A|Cross-references: EMBL:X17208; NID:g18136; PIDD:CAA5080.1; PID:g18137
A|Genetics:
A|Gene: cgcr-4

Query March          42.6%; Score 89; DB 2; Length 265;
Match Local Similarity 68.8%; Pred. No. 1,1;
Matches 22; Conservative 3; Mismatches 0; Gaps 0; Indels 7;

QY      3 AAAAAAAAAAAAAEAAAAAAAAAKKKKKKXK 34
        |||||:|||||::|::|::|::|::|::|
DB      176 AAEEARAAAAAAEAARAAAARAAAARAAKKAADK 207

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 / Search time 10.1408 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-4
Perfect score: 209
Sequence: 1 CEAAAAAEEAAAAAEEAAAAE.....KKKKKKKKKKKKKKKKKK 45

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	49.3	248	H1_PARAN	P02256 parachinus
2	101.5	48.6	442	YO11_SCHPO	O9Y718 scchirosach
3	96	45.9	433	ZU01_YEAST	P33527 saccharomyc
4	93	44.5	332	PA9A_HUMAN	O81201 homo sapien
5	90	43.1	153	H2B4_CHLRE	P53347 chlamydomon
6	89	42.6	207	H11_CABEL	P10771 caenorhabdi
7	86	41.1	347	TOLA_PSEAE	P50600 pseudomonas
8	85	40.7	102	RL12_METAE	P54346 methanococc
9	85	40.7	153	H2B3_CHLRE	P54345 chlamydomon
10	85	40.7	240	H12_VOLCA	O08865 volvox cart
11	84	40.2	441	PHPA_PLACH	O02752 plasmidium
12	84	40.2	726	BRD3_HUMAN	O15059 homo sapien
13	83	39.7	107	RL12_PYRAB	O9YX66 pyrococcus
14	83	39.7	320	YD33_YEAST	O12117 saccharomyc
15	83	39.7	421	TOLA_BCOLI	P19934 escherichia
16	82	39.2	153	H2B1_CHLRE	P50565 chlamydomon
17	82	39.2	474	CHP5_SCHPO	O14007 scchirosach
18	81	38.8	108	RL12_PYRAB	O57705 pyrococcus
19	81	38.8	487	CHP5_ASPRU	O43102 aspergillus
20	80.5	38.5	534	NOP5_RAT	O94286 rattus norv
21	80.5	38.5	1002	IF2P_YEAST	P39730 saccharomyc
22	80	38.3	111	RL12_MERPE	O9Y9W9 aeopytrum p
23	80	38.3	508	H1E_CHITE	P40278 chitronomus
24	80	38.3	508	NO60_DRORE	O44081 dirosophila
25	79.5	38.0	156	H2B2_CHLRE	P54345 chlamydomon
26	79.5	38.0	511	NOP5_YEAST	O12499 saccharomyc
27	79.5	38.0	2231	SENI_YEAST	O00416 saccharomyc
28	79	37.8	109	RLA1_TRYCR	P26643 trypanosoma
29	79	37.8	171	H1_ECHCR	P02257 echinolaipa
30	79	37.6	260	H11_VOLCA	O08864 volvox cart
31	78.5	37.6	1391	MS72_DROHY	O08696 dirosophila
32	78	37.3	134	H1_EDEPU	P31103 euploes eu
33	78	37.3	210	H1_LYRPI	P06144 lytechinus

ALIGNMENTS

RESULT 1	ID	H1_PARAN	STANDARD	PRT	248 AA.
AC	P02256				
DT	21-JUL-1986	(Rel. 01, Created)			
DT	21-JUL-1986	(Rel. 01, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	Histone H1, gonadal.				
OS	Parechinus angulosus (Angulate sea urchin).				
OC	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;				
OC	Echinoidae; Euechinoidae; Echinacea; Echinidae; Echinidae;				
OC	Parechinus.				
OX	NCBI_TaxID=7658;				
RM	[1]				
RP	SEQUENCE OF 1-84.				
RX	MEDLINE=80156831; PubMed=6767609;				
RA	Strickland W.N., Strickland M., de Groot P.C., von Holt C.,				
RA	Wittmann-Liebold B.;				
RT	"The primary structure of histone H1 from sperm of the sea urchin				
RT	Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the				
RT	protein and the sequence of amino acids in the four N-terminal				
RT	cyanogen bromide peptides.";				
RL	Eur. J. Biochem. 104:559-566(1980).				
RM	[2]				
RP	SEQUENCE OF 80-248.				
RX	MEDLINE=80156832; PubMed=7363905;				
RA	Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,				
RA	Wittmann-Liebold B.;				
RT	"The primary structure of histone H1 from sperm of the sea urchin				
RT	Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and				
RT	the entire primary structure.";				
RL	Eur. J. Biochem. 104:567-578(1980).				
CC	-1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF				
CC	-1- NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.				
CC	-1- SUBCELLULAR LOCATION: Nuclear.				
CC	-1- TISSUE SPECIFICITY: SPERM.				
CC	-1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.				
DR	PIR: A91090; HSURIP.				
DR	HSSP; P02259; IHST.				
DR	InterPro; IPR005818; Histone_H1/H5.				
DR	InterPro; IPR005819; Histone_H5.				
DR	InterPro; IPR003216; LinkerHist_N.				
DR	Pfam; PF00538; linker histone; I.				
DR	PRINTS; PR00624; HISTONEH5.				
DR	ProDom; PD000373; LinkerHist_N; 1.				
DR	SMART; SM00526; H1S; 1.				
KW	Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;				
FT	Sperm.				
FT	VARIANT				
SO	SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;				
QY	Query Match	49.3%;	Score 103;	DB 1;	Length 248;
	Best Local Similarity	59.5%;	Pred. No. 0.019;		
	Matches 25; Conservative	4;	Mismatches 13;	Gaps 0;	
	2 EAAAAAEEAAAAAEEAAAAE.....KKKKKKKKKKKKKKKKKK 43				

DB 146 KAAAKKAAALAKKAAAKKAAAKKAAAKKAAAKKAAAKK 187

RESULT 2

Y011 SCHPO STANDARD; PRT; 442 AA.
AC 09y18: 014347;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical J-domain protein C1778.01c in chromosome II.
GN SPBC1778.01C OR SPBC30D10.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagele K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynolprez B.,
RA Wellens J., Vantervelds E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Meisel D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Egger P., Zimmermann W., Medler H., Wandut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cernuti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; AL049489; CAB39796.1; -
DR EMBL; Z97992; CAB10796.1; -
DR PIR; T39683; T39683.
DR HSSP; P25685; 1HDJ.
DR GeneDB; SPombe; SPBC1778.01C; -
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS00636; DnaJ_2; 1.
DR PROSITE; PS00636; DnaJ_2; 1.
KM Hypothetical protein; Chapone.
FT DOMAIN 97 167 J-DOMAIN.
FT DOMAIN 294 342 ALA/LYS-RICH.
SQ SEQUENCE 442 AA; 50209 MW; F4EC924871B7118B CRC64;

Query Match 48.6%; Score 101.5; DB 1; Length 442;
Best Local Similarity 56.0%; Pred. No. 0.042;
Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;

RESULT 3

Y011 YEAST STANDARD; PRT; 433 AA.
AC P2527;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zuo1n.
GN ZUO1 OR YGR285C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=20B-12;
RX MEDLINE=93010971; PubMed=1396572;
RA Zhang S., Lockshin C., Herbert A., Winter E., Rich A.,
RA Zuo1n, a putative Z-DNA binding protein in Saccharomyces
RT cerevisiae.";
RL EMBL J. 11:3787-3796(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=6288C / FY1679;
RX MEDLINE=97245295; PubMed=9090054;
RA Volckaert G., Vest M., Robben J.;
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT the MLI1 locus reveals 15 complete open reading frames, including
RT ZUO1, BGJ2 and BJO2 genes and an ABC transporter gene.";
RL Yeast 13:251-259(1997).
CC -1- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN
CC CHROMOSOME ORGANIZATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; X63612; CAA45156.1; -
DR EMBL; Z73070; CAA97317.1; -
DR PIR; S25194; S25194.
DR HSSP; P08622; 1BQZ.
DR SGD; S0003517; ZUO1.
DR GO; GO:0005829; C:cytosol; NAS.
DR GO; GO:0005840; C:ribosome; NAS.
DR GO; GO:0003754; F:chaperone activity; NAS.
DR GO; GO:0006457; P:protein folding; NAS.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS00636; DnaJ_2; 1.
DR PROSITE; PS00636; DnaJ_2; 1.
KM Chaperone; DNA-binding; Nuclear protein.
FT DOMAIN 98 170 J-DOMAIN.
FT DOMAIN 306 357 ALA/LYS-RICH.
SQ SEQUENCE 433 AA; 49019 MW; 0AA76B1CD13C7DAB CRC64;
Query Match 45.9%; Score 96; DB 1; Length 433;

CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MICELLANEOUS: THERE ARE 2 FORMS OF H1 IN THIS NEMATODE: H1.1 IS
 CC THE MAJOR FORM.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53277; CAA37372.1; -.
 DR HSSP: P02259; IHST.
 DR InterPro: IPR005818; Histone_H1/H5.
 DR InterPro: IPR005819; Histone_H5.
 DR InterPro: IPR003216; Linkerhist_N.
 DR Pfam: PF00538; linker histone; 1.
 DR PRINTS: PR00624; HISTONEH5.
 DR ProDom: PD000373; Linkerhist_N; 1.
 DR SMART: SM00525; H15; 1.
 DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
 KM Acetylation.
 FT INIT MET 0 0
 FT MOD RES 1 1 ACETYLATION (PROBABLE).
 FT DOMAIN 36 112 GLOBULAR.
 FT CONFLICT 83 83 H -> L (IN REF. 2).
 FT CONFLICT 100 100 R -> K (IN REF. 2).
 SQ SEQUENCE 207 AA; 21314 MW; 7802EA9AA4F36F6F CRC64;

Query Match 42.6%; Score 89; DB 1; Length 207;
 Best Local Similarity 53.5%; Pred. No. 0.23; 16; Indels 0; Gaps 0;
 Matches 23; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 2 EAA 44
 DB 112 EKAAPAAKKPAAPAAKKPAAPAAKKPAAPAAKKPAAPAAKKPAAPAAKKPAAPAAKK 154

RESULT 7
 ID TOLA_PSEAE STANDARD; PRT; 347 AA.
 AC P50600;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE TOLA protein.
 GN TOLA OR PA0971.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAO;
 RX MEDLINE=97113525; PubMed=8955385;
 RA Dennis J.J., Lafontaine E.R., Sokol P.A.;
 RT "Identification and characterization of the tolORA genes of
 RT Pseudomonas aeruginosa.";
 RL J. Bacteriol. 178:7059-7068(1996).
 RN [2]
 RP REVISIONS TO N-TERMINUS.
 RA Duan X., Sokol P.A.;
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAO1;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Britman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goley L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
 RT opportunistic pathogen.";
 RL Nature 406:599-964(2000).
 CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
 CC (potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U39558; AAC44660.2; -.
 DR EMBL: AB004530; AAG04360.1; -.
 DR PIR: E83525; E83525.
 DR InterPro: IPR006260; TonB_C.
 DR TIGRFAMs: TIGR01352; tonB_Cterm; 1.
 KM Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
 KM Complete proteome.
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 37 POTENTIAL.
 FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).
 FT DOMAIN 209 216 POLY-ALA.
 SQ SEQUENCE 347 AA; 37935 MW; EEDDA804AA095945 CRC64;

Query Match 41.1%; Score 86; DB 1; Length 347;
 Best Local Similarity 51.2%; Pred. No. 0.61; 16; Indels 0; Gaps 0;
 Matches 22; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 2 EAA 44
 DB 124 EAQAAAEAKKDAEAKKAFAKAAEQAKQADYAKRAEDAEAKK 166

RESULT 8
 ID RL12_METUA STANDARD; PRT; 102 AA.
 AC P54048;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 50S ribosomal protein L12P.
 GN RPL12P OR M50508.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8689087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton W.D., Roberts K.M., Huret M.A., Kaine B.P., Borodovsky M.,
 RA Kleink H.-P., Frazer C.M., Smith H.O., Weese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
 CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
 CC ACCURATE TRANSLATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DC EMBL; U67500; AAB98498.1; -.
DR PIR; D64363; D64363.
DR TIGR; M00508; -.
DR InterPro; IPR001813; 60S ribosomal.
DR Pfam; PF00428; 60S ribosomal; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 102 AA; 10363 MW; 35306FFDE967C52C CRC64;

Query Match 40.7%; Score 85; DB 1; Length 102;
Best Local Similarity 52.6%; Pred. No. 0.26;
Matches 20; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAEMAAAEAAAAAAAAAKKKKKKKKKKK 39
Db 52 EAINAAAPVAAAPAAAPAAAPAAAEKKEEKKEEKKEE 89

RESULT 9
H2B3 CHURE STANDARD; PRT; 153 AA.
AC P54346;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DS Histone H2B-II1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonae.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Mueller K., Lindauer A., Park P.B., Cornelius T.,
RT Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas
RT histone genes reveal features linking plant and animal genes.";
RL Curr. Genet. 28:333-345(1995).
CC -1- SUBUNIT: The nucleosome is an octamer containing two molecules
CC each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC bp of DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the histone H2B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DC EMBL; U16725; AAA98450.1; -.
DR PIR; S59587; S59587.
DR InterPro; IPR004822; Histone core.
DR InterPro; IPR000558; Histone_H2B.
DR Pfam; PF00125; histone; 1.
DR PRINTS; PR00621; HISTONEH2B.
DR PRODOM; PD000497; Histone_H2B; 1.
DR SMART; SM00427; H2B; 1.
DR PROSITE; PS00357; HISTONE_H2B; 1.
KW Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW Multigene family.
SQ SEQUENCE 153 AA; 16557 MW; 2092413E04E1F49C CRC64;

Query Match 40.7%; Score 85; DB 1; Length 153;

```

```

Best Local Similarity 49.0%; Pred. No. 0.37;
Matches 24; Conservative 2; Mismatches 17; Indels 6; Gaps 1;
Cy 3 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAKXXXXXXXXXXXXX-----KXXXXXXXXX 45
Db 12 AEAAGAEAPAKAEAKPKAEKAGKAKKAKKAEKAKKAEKAKKAEKAKKAEKAKKAEKAKK 60

RESULT 10
ID H12_VOLCA STANDARD; PRT; 240 AA.
AC Q08865;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H1-II.
GN H1-II.
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=f. Nagariensis / HK10;
RX MEDLINE=93328125; PubMed=8335260;
Rt Lindauer A., Mueller K., Schmitt R.;
Rt "Two histone H1-encoding genes of the green alga Volvox carteri with
Rt features intermediate between plant and animal genes.";
Rt Gene 129:59-68(1993).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS RESTRICTED TO EMERYOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07947; AAA34246.1; -
DR PIR; JN0748; JN0748.
DR InterPro; IPR005818; Histone H1/H5.
DR InterPro; IPR005819; Histone H5.
DR InterPro; IPR003216; linkerhist_N.
DR Pfam; PF00538; linker_histone_1.
DR PRINTS; PR00624; HISTONEH.
DR ProDom; PD000373; linkerhist_N; 1.
DR SMART; SM00526; H1f. 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Repeat.
FT INIT_MET 0
FT DOMAIN 98 BY SIMILARITY.
FT DNA_BIND 182 185 GLOBULAR.
FT DNA_BIND 202 205 POTENTIAL.
FT DNA_BIND 210 213 POTENTIAL.
FT DOMAIN 110 216 8 X 6 AA REPEATS OF P-K-K-A-[AK]-A.
FT REPEAT 110 115 1.
FT REPEAT 116 121 2.
FT REPEAT 122 127 3.
FT REPEAT 183 188 4.
FT REPEAT 189 194 5.
FT REPEAT 195 200 6.
FT REPEAT 203 208 7.
FT REPEAT 211 216 8.
SQ SEQUENCE 240 AA; 25072 MW; 01AA389E08F421BD CRC64;
Query Match 40.7%; Score 85; DB 1; Length 240;
Best Local Similarity 50.0%; Pred. NO. 0.54;
Matches 23; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

```

```

QY      2 EAAAAAEAAAAAEAAAAAEEEEAAAAKKKKKKKKKKKKKKKK--KKKKKKK 45
          : ||| : ||| : ||| ||| ||| : ||| : ||| : |||
Db      99 KAKAAAKPKAAPKCAAAAPKCAAAAPKCAKAPKKEGKEKAAVVKPKSEKK 144

```

ID	PHPA_PLACH	STANDARD;	PRT;	441 AA.
AC	002752;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-JUN-1994 (Rel. 29, Last annotation update)			
DE	Acidic phosphoprotein precursor (50 kDa antigen).			
GN	PCMA1.			
OS	Plasmodium chabaudi.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.			
OX	NCBI_TaxID=5825;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IP-PC1;			
RX	MEDLINE=93116806; PubMed=1475002;			
RA	Deleensnijder W., Praemsetti P., Tungradrakul S., Hendrix D.,			
RA	Hamer-Casterman C., Hamers R.;			
RT	"Structure of a Plasmodium chabaudi acidic phosphoprotein that is			
RT	associated with the host erythrocyte membrane.";			
RL	Mol. Biochem. Parasitol. 56:59-68(1992).			
CC	-1- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES			
CC	THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE			
CC	PARASITE. ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.			
CC	-1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE			
CC	CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.			
CC	-1- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE			
CC	THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M95789; AAA29732.1; -.			
DR	PIR; A48455; A46455.			
KW	Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.			
FT	1	15		
FT	CHAIN	441		
FT	DOMAIN	186	313	
FT	REPEAT	186	193	16 X 8 AA TANDEM REPEATS.
FT	REPEAT	194	201	1-1.
FT	REPEAT	202	209	1-2.
FT	REPEAT	210	217	1-3.
FT	REPEAT	218	225	1-4.
FT	REPEAT	226	233	1-5.
FT	REPEAT	232	241	1-6.
FT	REPEAT	242	249	1-7.
FT	REPEAT	250	257	1-8.
FT	REPEAT	258	265	1-9.
FT	REPEAT	266	273	1-10.
FT	REPEAT	274	281	1-11.
FT	REPEAT	282	289	1-12.
FT	REPEAT	290	297	1-13.
FT	REPEAT	298	305	1-14.
FT	REPEAT	306	313	1-15.
FT	REPEAT	353	370	1-16.
FT	DOMAIN	353	360	2 X 9 AA TANDEM REPEATS.
FT	REPEAT	361	368	2-1.
FT	REPEAT	371	417	2-2.
FT	DOMAIN	21	21	LYS-RICH (BASIC).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	441 AA;	49708 MW;	N-LINKED (GLCNAC. . .) (POTENTIAL).
QO	SEQUENCE			DB55B3E795E7B5 CRC64;

Query Match	40.2%	Score 84	DB 1	Length 441
Best Local Similarity	80.0%	Pred. No. 1.1		
Matches	16	Conservative	4	Mismatches 0; Indels 0; Gaps 0;
Qy	26	XXXXXXXXXXXXXXXXXXXX	45	
Db	394	XXXXXXXXXXXXXXXXXXXX	413	

ID	BRD3_HUMAN	STANDARD;	PRT;	726 AA.
AC	Q15059; Q92645;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bromodomain-containing protein 3 (RING3-like protein).			
GN	BRD3 OR RING3L OR KIAA0043.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxId=9606;			
PN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
RX	MEDLINE=96051398; PubMed=7584044;			
RA	Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,			
RA	Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;			
RT	"Prediction of the coding sequences of unidentified human genes. II.			
RT	The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by			
RT	analysis of cDNA clones from human cell line KG-1."			
RL	DNA Res. 1:223-229(1994).			
RL	[2]			
RP	SEQUENCE OF 363-726 FROM N.A.			
RX	MEDLINE=98038990; PubMed=9373153;			
RA	Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;			
RT	"Chromosomal localization, gene structure and transcription pattern of			
RT	the ORFX gene, a homologue of the MHC-linked RING3 gene."			
RL	Gene 200:177-183(1997).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (potential).			
CC	-1- TISSUE SPECIFICITY: Ubiquitous.			
CC	-1- SIMILARITY: Contains 2 bromodomains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D26362; BAA05393.1; -			
DR	EMBL; Z81330; CAB03630.1; -			
DR	HSSP; Q92831.1B91.			
DR	Genew; HGNC:1104; BRD3.			
DR	MTM; 601541; -			
DR	GO; GO:0005634; C:nucleus; NAS.			
DR	InterPro; IPR001487; Bromodomain.			
DR	Pfam; PF00439; bromodomain; 2.			
DR	PRINTS; PR00503; BROMODOMAIN.			
DR	SMART; SM00287; BROMO. 2.			
DR	PROSITE; PS00633; BROMODOMAIN 1; 2.			
DR	PROSITE; PS50014; BROMODOMAIN 2; 2.			
KW	Bromodomain; Repeat; Nuclear protein.			
FT	DOMAIN 56 115 BROMODOMAIN 1.			
FT	DOMAIN 326 398 BROMODOMAIN 2.			
FT	DOMAIN 487 555 LYS-RICH.			
FT	DOMAIN 676 725 SER-RICH.			
FT	CONFILCT 465 466 EL -> DV (IN REF. 2).			
SO	SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;			
Query Match	40.2%;	Score 84;	DB 1;	Length 726;
Best Local Similarity	54.3%;	Pred. No. 1.7;		

Matches 19; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 11 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
DB 472 AVHEQLAALSQLPVNPKKKKKKKKKKKKKK 506

RESULT 13

ID_RL12_PYRAB STANDARD; PRT; 107 AA.

AC Q9UXS6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 50S ribosomal protein L12P
GN RPL12P OR PYRAB17820 OR PAB1168.
OS Pyrococcus abyssi.
OC Archaeae; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;

[1] SEQUENCE FROM N.A.

RP STRAIN=GES / Orsay;
RX PubMed:12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Hellig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost U., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).

CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AJ248288; CAB50687.1; -.

DR PIR; A75031; A75031.

DR InterPro; IPR001813; 60S_ribosomal.

DR Pfam; PF00428; 60S_ribosomal; 1.

KW Ribosomal protein; Complete proteome.

SEQUENCE 107 AA; 11300 MW; CD423930BAD4ED CRC64;

Query Match 39.7%; Score 83; DB 1; Length 107;
Best Local Similarity 44.2%; Pred. No. 0.39;
Matches 19; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
DB 55 EGAAMPVAVAAPAAPAPAEKKEKEKEKEKEEVSSE 97

RESULT 14

ID_YD33_YEAST STANDARD; PRT; 320 AA.

AC Q12117;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 36.2 kDa protein in RAD28-LYS14 intergenic region.
GN YDR033W OR YD673.03
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RX [1]

RP SEQUENCE FROM N.A.

RA Arnold W., Becker A., Jaeger W., Kuester H., Nussbaumer B.;
RU Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RU Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL OPSIN FAMILY. HSP30
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z74329; CAA98855.1; -.

DR EMBL; Z68196; CAA92370.1; -.

DR PIR; S61586; S61586.

DR SGP; S0002440; YDR033W.

DR GO; GO:0005886; C:plasma membrane; IDA.

DR InterPro; IPR001425; Bac_rhodopsin.

DR Pfam; PF01036; Bac_rhodopsin; 1.

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 35 55 POTENTIAL.

FT TRANSMEM 63 83 POTENTIAL.

FT TRANSMEM 117 137 POTENTIAL.

FT TRANSMEM 142 162 POTENTIAL.

FT TRANSMEM 168 188 POTENTIAL.

FT TRANSMEM 205 225 POTENTIAL.

FT TRANSMEM 239 259 POTENTIAL.

FT DOMAIN 300 318 LYS-RICH.

SEQUENCE 320 AA; 36190 MW; 4311F64B6AA209F CRC64;

Query Match 39.7%; Score 83; DB 1; Length 320;
Best Local Similarity 48.8%; Pred. No. 1;
Matches 21; Conservative 4; Mismatches 14; Indels 4; Gaps 1;

RESULT 15

ID_TOLA_ECOLI STANDARD; PRT; 421 AA.

AC P19934;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TOLA protein
GN TOLA OR CIM OR EXCC OR LXY OR B0739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JM105;
RX MEDLINE=90078104; PubMed=2687247;
RA Levensgood S.K., Webster R.E.;
RT "Nucleotide sequences of the tola and tolB genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=X12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;

RESULT 2
 Q8INH9 PRELIMINARY; PRT; 2347 AA.
 ID Q8INH9;
 AC Q8INH9;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE CG7518-PB.
 GN CG7518.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OC NCBI_taxid=7227;
 (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutcliffe G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bortova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris N.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Ibegwan C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Mlisina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Teeter C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster.";
 Science 287:2185-2195(2000).
 (12)
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.M., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Ferreira S., Frise E., Galie R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J.M., Paragay V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Peltman G.S., Puti V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitek R., Teeter C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (3)
 RP SEQUENCE FROM N.A.
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochick S.B., Smith C.D.,
 RA Tudy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RA "Annotation of Drosophila melanogaster genome.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (4)
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (5)
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003698; AA014338.1; -
 SQ SEQUENCE 2347 AA; 257013 MW; 23BF5FC5FCACAE64 CRC64;
 Query Match 56.5%; Score 118; DB 5; Length 2347;
 Best Local Similarity 75.7%; Pred. No. 0.021;
 Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 3 AAAAAAAAAAAAAAAAAAKKKKKKKKKKK 39
 DB 1374 AAAAAAAAAAAAAAAAAEQAQKKNKKQAKK 1410
 RESULT 3
 Q9VG05 PRELIMINARY; PRT; 2451 AA.
 ID Q9VG05;
 AC Q9VG05;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE CG7518 protein.
 GN CG7518.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OC NCBI_taxid=7227;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutcliffe G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bortova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris N.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Ibegwan C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Wuzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussekem D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltschbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.,
RT "Type genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003698; AAF54888.2; -.
DR FlyBase; FBgn0038108; CG7518.
DR InterPro; IPR001005; MYB_DNA_binding.
DR PROSITE; PS00037; MYB.1.1.
SQ SEQUENCE 2451 AA; 266959 MW; 088A2293F7481E2 CRC64;

Query Match 56.5%; Score 118; DB 5; Length 2451;
Best Local Similarity 75.7%; Pred. No. 0.022;
Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAKKKKKKKKKKK 39
DB 13/4 AAAAAAAAAAAAAAAAAEOKAKLKNKKOKAK 1410

RESULT 4

ID Q8T2U7 PRELIMINARY; PRT; 791 AA.

AC Q8T2U7; (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 92.4 kDa protein.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4;
RA Gieseckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Patra G., April J.F., Gulgo R., Kumpf K.,
RA Tungsang B., Cox B., Quail M.A., Platzner M., Rosenthal A., Noegel A.A.,
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115574; AAL92183.1; -.
DR InterPro; IPR005033; YEATS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF03366; YEATS.1.
DR SMART; SMO0355; ZNF_C2H2; 1.
KW Hypothetical protein.
SQ SEQUENCE 791 AA; 92375 MW; D6CCB6DEC92352C CRC64;

Query Match 51.2%; Score 107; DB 5; Length 791;
Best Local Similarity 56.8%; Pred. No. 0.07;
Matches 25; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAAAAAAAAAKKKKKKKKKKK 45
DB 746 EKEIETEMIGKEIEAEIEIEKEKKKKKKKKKKKKKKKKKK 789

RESULT 5

ID O3S807 PRELIMINARY; PRT; 129 AA.

AC O3S807; (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE MICROVASCULAR endothelial differentiation protein 2.
GN MD02.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Epithelium;
RX MEDLINE=96172708; PubMed=9511718;
RA Proles F., Loser B., Marx M.,
RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
RT species, during in vitro angiogenesis.";
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL; Y08769; CAA70022.1; -.
DR InterPro; IPR00719; Prot_kinase.
DR Pfam; PF00069; Kinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272B8E2EDB4 CRC64;

Query Match 50.2%; Score 105; DB 11; Length 129;
Best Local Similarity 95.5%; Pred. No. 0.02;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAAAAAAAAAAAAAAAAAKKKKKKKKKKK 45
DB 83 ASKKKKKKKKKKKKKKKKKKKKKKKKKK 104

RESULT 6

ID Q9H6Q7 PRELIMINARY; PRT; 720 AA.

AC Q9H6Q7; (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegaki T., Sugano S.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025632; BAB15196.1; -.
KW Hypothetical protein.
FT NON TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953DOB CRC64;

Query Match 50.2%; Score 105; DB 4; Length 720;
Best Local Similarity 61.8%; Pred. No. 0.095;
Matches 21; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 12 AAAAAAAAAAAAAAAAAAKKKKKKKKKKK 45
DB 678 AKKSITNSDIVISIKKKKKKKKKKKKKKKKKKKKKKKKK 711

RESULT 7

ID O8CGI8 PRELIMINARY; PRT; 658 AA.

AC O8CGI8; (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FVB/N;
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC035210, AAH35210.1; -.
KW Hypothetical protein.
FT NON TER 658
SQ SEQUENCE 658 AA; 73538 MW; 270E821463535FF9 CRC64;

Query Match 49.8%; Score 104; DB 11; Length 658;
Best Local Similarity 64.7%; Pred. No. 0.11;
Matches 22; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 11 AAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKKKKKKK 44
Db 625 ATSTTCTATVQAASSKKKKKKKKKKKKKKKKKKKK 658

RESULT 8
Q8LOP6 PRELIMINARY; PRT; 113 AA.
ID Q8LOP6
AC Q8LOP6
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE OJ117_G01.13 protein.
GN OJ117_G01.13
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RC Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
RT clone:OJ117_G01."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AP003374, BAB93330.1; -.
DR Gramene; Q8LOP6; -.
SQ SEQUENCE 113 AA; 13660 MW; 597DB0EBE22AA3EF CRC64;

Query Match 49.3%; Score 103; DB 10; Length 113;
Best Local Similarity 73.3%; Pred. No. 0.027; 7; Indels 0; Gaps 0;
Matches 22; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 16 AAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKKKKKKK 45
Db 3 ATSLHHKKKKKKKKKKKKKKKKKKKKKKKKKKKK 32

RESULT 9
Q9LL82 PRELIMINARY; PRT; 467 AA.
ID Q9LL82
AC Q9LL82
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Cbf5p.
OS Euglena gracilis.
OC Eukaryota; Euklenozoa; Euglenida; Euglenales; Euglena.
OX NCBI_TaxID=3039;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=streptomycin-bleached strain;
RC MEDLINE=20330353; PubMed=10871366;
RX Watanabe Y., Gray M.W.;
RT "Evolutionary appearance of genes encoding proteins associated with

```

```

RT box H/ACA snoRNAs: Cbf5p in Engelma girecillii, an early diverging
RT eukaryote, and candidate Garp and Nop10p homologs in
RT archaeobacteria." ;
RL Nucleic Acids Res. 28:2342-2352(2000) .
DR EMBL; AF234319; AAF77119.1; -.
DR InterPro; IPR004802; CBP5.
DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; TruB_N.
DR InterPro; IPR004521; Unchar_dom_2.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; TruB_N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGRFAMs; TIGR00425; CBP5; 1.
DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
SO SEQUENCE 467 AA; 52385 MW; 741089B6507BA7B CRC64;

Query Match 48.8%; Score 102; DB 10; Length 467;
Best Local Similarity 55.8%; Pred. No. 0.12;
Matches 24; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAK 44
Db 420 EELVAAEAAKKREREAAGEDEKDAKKAKKEKKEKKEKKEK 462

RESULT 10
O9LKR2 PRELIMINARY; PRT; 517 AA.
AC O9LKR2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 59.7 kDa protein.
GN T70N10_250.
OS Arabidopsis thaliana (Mouse-ear cress) .
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxId=3702;
CK [1]
RN RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G.,
RA Rudd S., Lemke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RV [2]
RN RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353032; CAB88307.1; -.
DR InterPro; IPR001810; F-box.
DR InterPro; IPR005566; FBD.
DR Pfam; PF00646; F-box; 1.
DR Pfam; PF00579; FBD; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PS50181; FBOX; 1.
KW Hypothetical protein.
SQ SEQUENCE 517 AA; 59689 MW; BC6D957D01F86770 CRC64;

Query Match 48.8%; Score 102; DB 10; Length 517;
Best Local Similarity 80.8%; Pred. No. 0.13;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 20 EAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAK 45
Db 440 ERVGGKKKKKKKKKKKKKKKKKKKKKKKKKKKK 465

RESULT 11
O14347 PRELIMINARY; PRT; 354 AA.
AC O14347;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
```

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Zucotin-like protein (Fragment).
 GN SPBC30D10.01.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 OX NCB1_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Beck A., Reinhardt R., Lyne M., Rajandream M.A., Barrell B.G.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 297992; CAB10796.1; -
 DR HSSP; P25685; 1HDJ.
 DR InterPro; IPR001623; DnaJ_N.
 DR Pfam; PF00226; DnaJ_1.
 DR SMART; SM00271; DnaJ_1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS0076; DnaJ_2; 1.
 KM Dna-binding.
 KW NON_TER
 SQ SEQUENCE 354 AA; 40290 MW; 6071B58A3B60F558 CRC64;
 Query Match 48.6%; Score 101.5; DB 3; Length 354;
 Best Local Similarity 56.0%; Pred. No. 0.1; Mismatches 12; Indels 7; Gaps 1;
 Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;

OY 2 EAAAAA-----EAAAAAEEEEAAAAAEEEEEEEEEEEEEEEEEEEE 44
 DB 221 EAAAAAGKKEEBERRAEBAAKASAAANAKAEDKKKKKKKKKKKKKK 270

RESULT 12

O95LV6 PRELIMINARY; PRT; 531 AA.
 AC O95LV6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical 61.4 kDa protein (Fragment).
 OS Macaca fascicularis (Crah eating macaque) (Cynomolgus monkey).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 CC Cercopithecinae; Macaca.
 OX NCB1_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
 RA Terato K., Sugano S.;
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
 RT libraries."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB071085; BAB64479.1; -
 KM Hypothetical protein.
 FT NON_TER 531
 SQ SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;
 Query Match 48.3%; Score 101; DB 6; Length 531;
 Best Local Similarity 95.2%; Pred. No. 0.16; Mismatches 0; Indels 0; Gaps 0;
 Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 25 AA 45
 DB 501 SKKK 521

RESULT 13
 O957D3 PRELIMINARY; PRT; 80 AA.
 AC O957D3;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical 9.4 kDa protein.
 GN OSUNBA0057L21.23.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhartioideae; Oryzaceae; Oryza.
 OX NCB1_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Bueli C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
 RA Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
 RA Teitelin T., Riggs P., Hsiao J., Zisemann V., Blunt S., Pal G.,
 RA Vanaken S.E., Utecherack T.R., Feldblum T.V., Kalb E., Quackenbush J.,
 RA Salzberg S.L., White O., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSUNBA0057L21 genomic sequence."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC087599; AAL79706.1; -
 DR Gramine; O957D3; -
 KW Hypothetical protein.
 SQ SEQUENCE 80 AA; 9362 MW; 0177C863133B21D8 CRC64;
 Query Match 48.1%; Score 100.5; DB 10; Length 80;
 Best Local Similarity 49.1%; Pred. No. 0.032; Mismatches 16; Indels 11; Gaps 1;
 Matches 27; Conservative 1; Mismatches 16; Indels 11; Gaps 1;

OY 2 EAAAAA-----EAAAAAEEEEAAAAAEEEEEEEEEEEEEEEEEEEE 45
 DB 15 EAAGGATGCGRGMWMPATATRVFLSVIHLDRKKKKKKKKKKKKKKKKKK 69

RESULT 14

O9N6FO PRELIMINARY; PRT; 55 AA.
 AC O9N6FO;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Similar to LOC201361.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Straubeberg R.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC030525; AAH30525.1; -
 SQ SEQUENCE 55 AA; 7251 MW; 0906032B284006BA CRC64;

Query Match 47.8%; Score 100; DB 4; Length 55;
 Best Local Similarity 100.0%; Pred. No. 0.026; Mismatches 0; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 KKK 45
 DB 21 KKK 40

RESULT 15

O9P529 PRELIMINARY; PRT; 128 AA.
 AC O9P529;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical 15.2 kDa protein.
 GN B24H17.160.
 OS Neurospora crassa.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariaceae; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohnselt J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356815; CAB92638.2; -
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 47.8%; Score 100; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 26 KKKKKKKKKKKKKKKKKKKKK 45
|||
Db 71 KKKKKKKKKKKKKKKKKKKKK 90

Search completed: January 30, 2004, 00:24:41
Job time : 46.4789 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 18.7981 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-5
Perfect score: 243
Sequence: 1 GLEFGAIFGFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table:
BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	134	55.1	1 HMIIVS2	hemagglutinin prec
2	134	55.1	1 HMIIVS3	hemagglutinin prec
3	134	55.1	1 HMIIV77	hemagglutinin prec
4	134	55.1	1 HMIIV80	hemagglutinin prec
5	134	55.1	1 HMIIV33	hemagglutinin prec
6	134	55.1	1 HMIIV89	hemagglutinin prec
7	134	55.1	1 HMIIV21	hemagglutinin prec
8	134	55.1	1 HMIIV98	hemagglutinin prec
9	134	55.1	1 HMIIV15	hemagglutinin prec
10	134	55.1	2 JQ1153	hemagglutinin prec
11	134	55.1	2 JQ1154	hemagglutinin prec
12	134	55.1	2 JQ1155	hemagglutinin prec
13	134	55.1	1 HMIIVH	hemagglutinin prec
14	134	55.1	1 HMIIVH	hemagglutinin prec
15	134	55.1	1 HMIIVH	hemagglutinin prec
16	134	55.1	1 HMIIVH	hemagglutinin prec
17	134	55.1	1 HMIIVH	hemagglutinin prec
18	134	55.1	1 HMIIVH	hemagglutinin prec
19	134	55.1	1 HMIIVH	hemagglutinin prec
20	133	54.7	1 HMIIV84	hemagglutinin prec
21	132	54.3	1 HMIIV81	hemagglutinin prec
22	132	54.3	1 HMIIV81	hemagglutinin prec
23	132	54.3	1 HMIIV81	hemagglutinin prec
24	132	54.3	1 HMIIV81	hemagglutinin prec
25	131	53.9	2 S38637	hemagglutinin - in
26	131	53.9	1 HMIIV86	hemagglutinin prec
27	131	53.9	1 HMIIV86	hemagglutinin prec
28	131	53.9	1 HMIIV86	hemagglutinin prec
29	131	53.9	1 HMIIV86	hemagglutinin prec

30	131	53.9	565 1 HMIIVS5	hemagglutinin prec
31	131	53.9	565 1 HMIIVS6	hemagglutinin prec
32	131	53.9	565 1 HMIIVS7	hemagglutinin prec
33	131	53.9	565 1 HMIIVS8	hemagglutinin prec
34	131	53.9	565 1 HMIIVS9	hemagglutinin prec
35	131	53.9	565 1 HMIIVS9	hemagglutinin prec
36	131	53.9	565 1 HMIIVS9	hemagglutinin prec
37	131	53.9	565 1 HMIIVS9	hemagglutinin prec
38	131	53.9	565 1 HMIIVS9	hemagglutinin prec
39	131	53.9	565 1 HMIIVS9	hemagglutinin prec
40	131	53.9	565 1 HMIIVS9	hemagglutinin prec
41	131	53.9	565 1 HMIIVS9	hemagglutinin prec
42	131	53.9	565 1 HMIIVS9	hemagglutinin prec
43	131	53.9	565 1 HMIIVS9	hemagglutinin prec
44	131	53.9	565 1 HMIIVS9	hemagglutinin prec
45	131	53.9	565 1 HMIIVS9	hemagglutinin prec

ALIGNMENTS

RESULT 1
HMIIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C:Species: influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: A29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: A29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19056; NID:g324208
A:Note: the sequence in Genbank entry FLAHABP, release 106, (PID:g324209) differs from the
C:Genetics:
A:Map position: segment 4
A:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein, hemagglutinin, homotrimer, lipoprotein, thiolester bond
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Product: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-486,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted
Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2 GLEFGAIFGFIENGWEGMIDGMYG 24
Db 330 GLEFGAIFGFIENGWEGMIDGMYG 352
RESULT 2
HMIIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C:Species: influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: B29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: B29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19057; NID:g324210
A:Note: the sequence in Genbank entry FLAHABP, release 106, (PID:g324211) differs from the
C:Genetics:
A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domains: transmembrane #status predicted <TM1>
 F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWTG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 3
 HMIIV77
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: A27813
 R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
 Virology 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: A27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16737; NID:G324081; PIDN:AAA43143.1; PID:G324082
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domains: transmembrane #status predicted <TM1>
 F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWTG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 4
 HMIIV80
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
 C:Accession: B27813
 R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
 Virology 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: B27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16738; NID:G324083
 A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domains: transmembrane #status predicted <TM1>
 F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWTG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 5
 HMIIV33
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: C27813
 R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
 Virology 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: C27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16739; NID:G324085; PIDN:AAA43145.1; PID:G324086
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domains: transmembrane #status predicted <TM1>
 F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWTG 24
 DB 330 GLFGAIAGFIENGWEGMIDGWTG 352

RESULT 6
 HMIIV89
 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
 N:Contains: hemagglutinin HA1; hemagglutinin HA2
 C:Species: influenza A virus
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: D27813
 R:Kida, H.; Kawaka, Y.; Naeve, C.W.; Webster, R.G.
 Virology 159, 109-119, 1987
 A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
 A:Reference number: A94363; PMID:87265458; PMID:2440178
 A:Accession: D27813
 A:Molecule type: genomic RNA
 A:Residues: 1-550 <KID>
 A:Cross-references: GB:M16740; NID:G324087; PIDN:AAA43146.1; PID:G324088
 C:Genetics:
 A:Map position: segment 4
 C:Superfamily: influenza virus hemagglutinin
 C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
 F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
 F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
 F:520-536/Domains: transmembrane #status predicted <TM1>

F/8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F/14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
 F/539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 7

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)

N/Contains: hemagglutinin HA1; hemagglutinin HA2

C/Species: influenza A virus

C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998

C/Accession: E27813

R/Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A/Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A/Reference number: A94363; MUID:87265458; PMID:2440178

A/Accession: E27813

A/Molecule type: genomic RNA

A/Residues: 1-550 <KID>

A/Cross-references: GB:M16741; NID:9324089

C/Genetics:

A/Map position: segment 4

C/Superfamily: influenza virus hemagglutinin

C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F/1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F/330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F/520-536/Domains: transmembrane #status predicted <TM1>

F/7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F/539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 8

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)

N/Contains: hemagglutinin HA1; hemagglutinin HA2

C/Species: influenza A virus

C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998

C/Accession: F27813

R/Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A/Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A/Reference number: A94363; MUID:87265458; PMID:2440178

A/Accession: F27813

A/Molecule type: genomic RNA

A/Residues: 1-550 <KID>

A/Cross-references: GB:M16742; NID:9324091

C/Genetics:

A/Map position: segment 4

C/Superfamily: influenza virus hemagglutinin

C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F/1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F/330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F/520-536/Domains: transmembrane #status predicted <TM1>

F/7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 9

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)

N/Contains: hemagglutinin HA1; hemagglutinin HA2

C/Species: influenza A virus

C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C/Accession: G27813

R/Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A/Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A/Reference number: A94363; MUID:87265458; PMID:2440178

A/Accession: G27813

A/Molecule type: genomic RNA

A/Residues: 1-550 <KID>

A/Cross-references: GB:M16743; NID:9324093; PIDN:AAA3149.1; PID:9324094

C/Genetics:

A/Map position: segment 4

C/Superfamily: influenza virus hemagglutinin

C/Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F/1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F/330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F/520-536/Domains: transmembrane #status predicted <TM1>

F/7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/14,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F/539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWYG 24
 Db 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 10

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)

N/Contains: hemagglutinin HA1; hemagglutinin HA2

C/Species: influenza A virus

C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C/Accession: J01153

R/Yaounde, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A/Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A/Reference number: J01153; MUID:91341491; PMID:1875195

A/Accession: J01153

A/Molecule type: genomic RNA

A/Residues: 1-550 <YAS>

A/Cross-references: GB:D00929; NID:9221279; PIDN:BA00769.1; PID:9221280

A/Note: the authors translated the codon GGG for residue 218 as Glu

A/Note: residues 528-532 are not shown in this publication

C/Superfamily: influenza virus hemagglutinin

C/Keywords: glycoprotein; hemagglutinin; homotrimer

F/1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F/330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F/8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 GLFGAIAGFIENGWEGMIDGWYG 24

```

Db      330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11
JQ1154 hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
N:Contains: hemagglutinin HA1, hemagglutinin HA2
C:Species: influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: JQ1154
R:Yaounda, J.; Shortridge, K.F.; Shintzu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A>Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: JQ1153; MUID:91341491; PMID:1875195
A:Accession: JQ1154
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00930; NID:9221273; PIDN:BA00770.1; PID:9221274
A>Note: the authors translated the codon GGG for residue 218 as G1U
A>Note: residues 528-532 are not shown in this publication
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      55.1%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
Db      330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 12
JQ1155 hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)
N:Contains: hemagglutinin HA1, hemagglutinin HA2
C:Species: influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: JQ1155
R:Yaounda, J.; Shortridge, K.F.; Shintzu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A>Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: JQ1153; MUID:91341491; PMID:1875195
A:Accession: JQ1155
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00931; NID:9221277; PIDN:BA00771.1; PID:9221278
A>Note: the authors translated the codon GGG for residue 218 as G1U, GCC for residue 538
A>Note: residues 528-532 are not shown in this publication
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      55.1%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
Db      330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 13
HMI1VH hemagglutinin precursor - influenza A virus
C:Species: influenza A virus
C>Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C:Accession: A93705; A93233; A04051; A93231; A94441

```

```

R.Both, G.W.; Sleight, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A>Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza A
A:Reference number: A93705; MUID:81053698; PMID:6253883
A:Accession: A93705
A:Molecule type: genomic RNA
A:Residues: 1-566 <BOT>
A:Cross-references: GB:V01103
A:Experimental source: strain A/NT/60/68/29C
A>Note: Human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/60
R:Dopheide, T.A.; Ward, C.W.
PBS Lett. 110, 181-183, 1980
A>Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A:Reference number: A91276; MUID:80179105; PMID:6768586
A:Contents: annotation; disulfide bonds
A:Contents: annotation; disulfide bonds
R:Geetha, M.U.; Bye, U.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A>Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from
A:Reference number: A93233; MUID:81030852; PMID:7421990
A:Accession: A93233
A:Molecule type: genomic RNA
A:Residues: 1-24, S', 26, D', 28-159, 'G', 161-197, 'I', 199-241, 'L', 243-249 <GRT>
A:Experimental source: strain X-31[H3]
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-244/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status experimental
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match      55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
Db      346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 14
HMI1VH hemagglutinin precursor - influenza A virus (strain A/ichi/2/68)
N:Contains: hemagglutinin HA1, hemagglutinin HA2
C:Species: influenza A virus
C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C:Accession: A93231; A04051
R:Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huybreoek, D.; Saman, E.; Fiers, W.
Nature 286, 771-776, 1980
A>Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A
A:Reference number: A93231; MUID:80254693; PMID:7402351
A:Accession: A93231
A:Molecule type: genomic RNA
A:Residues: 1-566 <VER>
A:Cross-references: GB:J02090; NID:9324131; PIDN:AA43178.1; PID:9324132
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-244/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match      55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
Db      346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 15

```


HMI17M

hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C:Date: 19-May-1994 #sequence revision 19-May-1994 #text_change 31-Mar-2000

C:Accession: A94441; A04051

R:Stleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.

in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, F

A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of C

A:Reference number: A94441

A:Accession: A94441

A:Molecule type: genomic RNA

A:Residues: 1-566 <SLR>

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-34/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;

Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIGFIENGWEGMIDGWYG 24

Db 346 GLFGAIGFIENGWEGMIDGWYG 368

Search completed: January 30, 2004, 00:26:22
Job time : 18.7981 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 9.91549 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-5
Perfect score: 243
Sequence: 1 CGFAGAIAGFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	550	1 HEMA_IADH1	P12582 Influenza a
2	134	55.1	550	1 HEMA_IADH2	P12583 Influenza a
3	134	55.1	550	1 HEMA_IADH3	P12584 Influenza a
4	134	55.1	550	1 HEMA_IADH4	P12585 Influenza a
5	134	55.1	550	1 HEMA_IADH5	P12586 Influenza a
6	134	55.1	550	1 HEMA_IADH6	P12587 Influenza a
7	134	55.1	550	1 HEMA_IADH7	P12588 Influenza a
8	134	55.1	550	1 HEMA_IADH8	P43257 Influenza a
9	134	55.1	550	1 HEMA_IADH9	P43258 Influenza a
10	134	55.1	550	1 HEMA_IADH10	P43260 Influenza a
11	134	55.1	550	1 HEMA_IADH11	P11133 Influenza a
12	134	55.1	550	1 HEMA_IADH12	P11134 Influenza a
13	134	55.1	550	1 HEMA_IADH13	P03437 Influenza a
14	134	55.1	550	1 HEMA_IADH14	P26134 Influenza a
15	134	55.1	550	1 HEMA_IADH15	P03442 Influenza a
16	134	55.1	550	1 HEMA_IADH16	P26138 Influenza a
17	134	55.1	550	1 HEMA_IADH17	P03449 Influenza a
18	134	55.1	550	1 HEMA_IADH18	P03439 Influenza a
19	134	55.1	550	1 HEMA_IADH19	P03436 Influenza a
20	134	55.1	550	1 HEMA_IADH20	P26094 Influenza a
21	134	55.1	550	1 HEMA_IADH21	P26101 Influenza a
22	133	54.7	561	1 HEMA_IADH22	P12581 Influenza a
23	133	54.7	561	1 HEMA_IADH23	P12439 Influenza a
24	132	54.3	565	1 HEMA_IADH24	P17000 Influenza a
25	132	54.3	565	1 HEMA_IADH25	P17002 Influenza a
26	132	54.3	566	1 HEMA_IADH26	P03440 Influenza a
27	132	54.3	566	1 HEMA_IADH27	P26139 Influenza a
28	132	54.3	567	1 HEMA_IADH28	P03435 Influenza a
29	131	53.9	560	1 HEMA_IADH29	P12589 Influenza a
30	131	53.9	560	1 HEMA_IADH30	P03458 Influenza a
31	131	53.9	565	1 HEMA_IADH31	P16994 Influenza a
32	131	53.9	565	1 HEMA_IADH32	P16995 Influenza a
33	131	53.9	565	1 HEMA_IADH33	P16999 Influenza a

34	131	53.9	565	1 HEMA_IADH34	P16996 Influenza a
35	131	53.9	565	1 HEMA_IADH35	P16997 Influenza a
36	131	53.9	565	1 HEMA_IADH36	P16998 Influenza a
37	131	53.9	565	1 HEMA_IADH37	P16999 Influenza a
38	131	53.9	565	1 HEMA_IADH38	P16999 Influenza a
39	131	53.9	565	1 HEMA_IADH39	P16999 Influenza a
40	131	53.9	565	1 HEMA_IADH40	P16999 Influenza a
41	131	53.9	565	1 HEMA_IADH41	P16999 Influenza a
42	131	53.9	565	1 HEMA_IADH42	P16999 Influenza a
43	131	53.9	565	1 HEMA_IADH43	P16999 Influenza a
44	131	53.9	565	1 HEMA_IADH44	P16999 Influenza a
45	131	53.9	565	1 HEMA_IADH45	P16999 Influenza a

ALIGNMENTS

```

RESULT 1
HEMA_IADH1 STANDARD; PRT; 550 AA.
AC P12582; Q84021; Q84022;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11357;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawasaka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119 (1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ac.uk).
CC
CC EMBL; P12582; AAA43143.1; -.
CC HSPD; P03437; 3HWG.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein.
CC NON TER 1
CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 550 AA; 61705 MW; 7E7ACFE716FC969A CRC64;
Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2,2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 2 GLFGAIAFIENGMEGMIDGWTG 24
 DB 330 GLFGAIAFIENGMEGMIDGWTG 352

RESULT 2

HEMA_IADH2 STANDARD: PRT: 550 AA.

AC P12583; Q84011; (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAI chain] (Fragment).
 GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/8/80).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OK NCBI_TaxID=11358;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:109-119(1987).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M16738; AAA43144.1; -
 DR CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 DR PIR; B27813; 2VIT.
 DR HSSP; P03437; 2VIT.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTIN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 8 550 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 137 137 K -> N (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFIENGMEGMIDGWTG 24
 DB 330 GLFGAIAFIENGMEGMIDGWTG 352

RESULT 3
 ID HEMA_IADH3 STANDARD: PRT: 550 AA.

AC P12584; Q84012; Q89793;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAI chain] (Fragment).
 GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OK NCBI_TaxID=11359;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:109-119(1987).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M16739; AAA43145.1; -
 DR CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 DR HSSP; P03437; 2VIT.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTIN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 8 550 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAFIENGMEGMIDGWTG 24
 DB 330 GLFGAIAFIENGMEGMIDGWTG 352

RESULT 4

HEMA_IADH4 STANDARD: PRT: 550 AA.

AC P12585; Q84013; Q84014;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAI chain] (Fragment).
 GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/7/82).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OK NCBI_TaxID=11360;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Nave C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
   ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M16740; AAA43146.1; -.
DR HSSP; P03437; 2V1U.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1 328
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBDB9D0 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDWYG 24
DB 330 GLFGAIGFIENGWEGMIDWYG 352

```

```

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M16741; AAA43147.1; -.
DR PIR; E27813; HMTV21.
DR HSSP; P03437; 2V1U.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1 328
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 7 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15BFC CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDWYG 24
DB 330 GLFGAIGFIENGWEGMIDWYG 352

```

```

RESULT 5
HEMA_IADH5 STANDARD; PRT; 550 AA.
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11361;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Nave C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
   ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

```

```

RESULT 6
HEMA_IADH6 STANDARD; PRT; 550 AA.
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Nave C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
   ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M16742; AAA43148.1; -

DR PIR: F27813; HMI98.

DR HSSP: P03437; 1HGJ.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR Prodom: PD000225; Hemagglutn; 1.

DR Envelope protein; Hemagglutinin; Glycoprotein.

KW CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT NON TER 1 328 HEMAGGLUTININ HAI CHAIN.

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).

FT SEQUENCE 550 AA; 61711 MW; 67BCDB5F44736CFE CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWYG 24
|||
Db 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 7

HEMA_IADH7 STANDARD; PRT; 550 AA.

ID HEMA_IADH7

AC P12588; Q84018; Q89470;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain; Hemagglutinin HAZ chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/10/85).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza A viruses; Influenzavirus A.

OK NCBI_TaxID=11363;

RN NCB1

RP SEQUENCE FROM N.A.

RA MEDLINE=87265458; PubMed=2440178;

RX Kida H., Kawacka Y., Naave C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild ducks."

RT Viology 159:109-119(1987).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M16743; AAA43149.1; -

DR HSSP: P03437; 3HMG.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR Prodom: PD000225; Hemagglutn; 1.

DR Envelope protein; Hemagglutinin; Glycoprotein.

FT NON TER 1 1

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 550 AA; 61761 MW; 68F81793281D53EB CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMIDGWYG 24
|||
Db 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 8

HEMA_IADHK STANDARD; PRT; 550 AA.

ID HEMA_IADHK

AC P43257;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain; Hemagglutinin HAZ chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hong Kong/7/75).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza A viruses; Influenzavirus A.

OK NCBI_TaxID=11364;

RN NCB1

RP SEQUENCE FROM N.A.

RA MEDLINE=91341491; PubMed=1875195;

RX Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;

RT "Molecular evidence for a role of domestic ducks in the introduction of avian H3 influenza viruses to pigs in southern China, where the

RT of avian H3 influenza viruses to pigs in southern China, where the

RT J. Gen. Virol. 72:2007-2010(1991).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D00929; BA00769.1; -

DR HSSP: P03437; 2VU.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR Prodom: PD000225; Hemagglutn; 1.

DR Envelope protein; Hemagglutinin; Glycoprotein.

KW CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT NON TER 1 328 HEMAGGLUTININ HAI CHAIN.

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 550 AA; 61549 MW; 864639B829FE1BA9 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2,2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GLFGAIAFGIENGWEGMIDWYG 24
 |||||
 Db 330 GLFGAIAFGIENGWEGMIDWYG 352

RESULT 9

HEMA_IADHL STANDARD; PRT; 550 AA.
 ID HEMA_IADHL
 AC P43258;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hong Kong/64/76).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CX NCBI_TaxID=45412;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yaoda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 RT of avian H3 influenza viruses to pigs in southern China, where the
 RT A/Hong Kong/68 (H3N2) strain emerged.";
 RL J. Gen. Virol. 72:2007-2010 (1991).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00931; BAA00771.1; -;
 DR HSSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 550 AA; 61718 MW; A351C56789B4BE9A CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2,2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GLFGAIAFGIENGWEGMIDWYG 24
 |||||
 Db 330 GLFGAIAFGIENGWEGMIDWYG 352

RESULT 10

HEMA_IAGHK STANDARD; PRT; 550 AA.
 ID HEMA_IAGHK
 AC P43260;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CX NCBI_TaxID=45414;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yaoda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 RT of avian H3 influenza viruses to pigs in southern China, where the
 RT A/Hong Kong/68 (H3N2) strain emerged.";
 RL J. Gen. Virol. 72:2007-2010 (1991).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00930; BAA00770.1; -;
 DR HSSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BAC02 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2,2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GLFGAIAFGIENGWEGMIDWYG 24
 |||||
 Db 330 GLFGAIAFGIENGWEGMIDWYG 352

RESULT 11

HEMA_IAGH2 STANDARD; PRT; 550 AA.
 ID HEMA_IAGH2
 AC P11133; Q84019; Q84020;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
 GN chain] (Fragment).
 HA.

OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CC NCB1_TaxID=11497;
 CC (1)
 CC SEQUENCE FROM N.A.
 CC MEDLINE=88101364; PubMed=3336940;
 CC Kida H., Shortridge K.F., Webster R.G.;
 CC "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 CC in China.";
 CC Virology 162:160-166(1988).
 CC
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M19057; AAA43212.1; -
 CC HSRP; P03437; 2V1U.
 CC InterPro: IPR001364; Hemagglutn.
 CC Pfam: PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; HEMAGGLUTN12.
 CC ProDom: PD000225; Hemagglutn; 1.
 CC Hemagglutinin; Envelope protein; Glycoprotein.
 CC
 CC NON_TER 1 1
 CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 CC
 CC FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
 CC FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;
 CC
 CC Query Match 55.1%; Score 134; DB 1; Length 550;
 CC Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 CC Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 CC |||||
 CC DB 330 GLFGAIAGFIENGWEGMIDGWYG 352
 CC
 CC RESULT 12
 CC HEMA_IASH3 STANDARD; PRT; 550 AA.
 CC ID HEMA_IASH3 Q84025; Q84026;
 CC AC P1134; Q84025; Q84026;
 CC DT 01-JUL-1988 (Rel. 11, Created)
 CC DT 01-JUL-1988 (Rel. 11, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
 CC chain] (Fragment).
 CC GN HA.
 CC OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CC NCB1_TaxID=11498;
 CC (1)
 CC SEQUENCE FROM N.A.
 CC MEDLINE=88101364; PubMed=3336940;
 CC Kida H., Shortridge K.F., Webster R.G.;
 CC "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 CC in China.";
 CC Virology 162:160-166(1988).
 CC

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M19056; AAA43211.1; ALT_TERM.
 CC HSRP; P03437; 2V1U.
 CC InterPro: IPR001364; Hemagglutn.
 CC Pfam: PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; HEMAGGLUTN12.
 CC ProDom: PD000225; Hemagglutn; 1.
 CC Hemagglutinin; Envelope protein; Glycoprotein.
 CC
 CC NON_TER 1 1
 CC CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 CC
 CC FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
 CC FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 550 AA; 61580 MW; 991F6D8BC02F24F2 CRC64;
 CC
 CC Query Match 55.1%; Score 134; DB 1; Length 550;
 CC Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 CC Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 CC |||||
 CC DB 330 GLFGAIAGFIENGWEGMIDGWYG 352
 CC
 CC RESULT 13
 CC HEMA_IAMIC STANDARD; PRT; 566 AA.
 CC ID HEMA_IAMIC P03437;
 CC AC P03437;
 CC DT 21-JUL-1986 (Rel. 01, Created)
 CC DT 21-JUL-1986 (Rel. 01, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 CC Hemagglutinin HA2 chain].
 CC GN HA.
 CC OS Influenza A virus (strain A/Aichi/2/68).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CC NCB1_TaxID=150147;
 CC (1)
 CC SEQUENCE FROM N.A.
 CC MEDLINE=80254693; PubMed=7402351;
 CC Verhoeven M., Fang R., Min Jou W., Devos R., Huylbrouck D.,
 CC Saman B., Fiers W.;
 CC "Antigenic drift between the haemagglutinin of the Hong Kong
 CC influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
 CC Nature 286:771-776(1980).
 CC (2)
 CC X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 CC MEDLINE=81123029; PubMed=7464906;
 CC Wilson I.A., Skehel J.J., Wiley D.C.;
 CC "Structure of the haemagglutinin membrane glycoprotein of influenza
 CC virus at 3-A resolution.";
 CC Nature 289:366-373(1981).
 CC (3)
 CC X-RAY CRYSTALLOGRAPHY.
 CC MEDLINE=88232903; PubMed=3374584;
 CC

FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIEGMEGMITDGYG 24

Db 346 GLFGAIAGFIEGMEGMITDGYG 368

Search completed: January 30, 2004, 00:20:46
Job time : 9.91549 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 45.446 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-5
Perfect score: 243
Sequence: 1 GCLFGALAGFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP_invertebrate:*
- 6: SP_mammal:*
- 7: SP_mmc:*
- 8: SP_organelle:*
- 9: SP_phage:*
- 10: SP_plant:*
- 11: SP_rodent:*
- 12: SP_virus:*
- 13: SP_vertebrate:*
- 14: SP_unclassified:*
- 15: SP_virus:*
- 16: SP_bacteriap:*
- 17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	384	12 Q8JRK63	Q8JRK63 Influenza a
2	134	55.1	566	12 Q98052	Q98052 Influenzavi
3	134	55.1	566	12 Q8U251	Q8U251 Influenza a
4	134	55.1	566	12 Q8OLN8	Q8OLN8 Influenza a
5	134	55.1	566	12 Q67132	Q67132 Influenza a
6	134	55.1	566	12 Q67135	Q67135 Influenzavi
7	134	55.1	566	12 Q8UXK3	Q8UXK3 Influenza a
8	134	55.1	566	12 Q91MA7	Q91MA7 Influenza a
9	134	55.1	566	12 Q9DHG0	Q9DHG0 Influenza a
10	134	55.1	566	12 Q910M5	Q910M5 Influenza a
11	134	55.1	566	12 Q67126	Q67126 Influenzavi
12	133	54.7	301	12 Q9DXK3	Q9DXK3 Influenza a
13	132	54.3	550	12 Q82498	Q82498 Influenzavi
14	132	54.3	550	12 Q82498	Q82498 Influenzavi
15	132	54.3	550	12 Q82753	Q82753 Influenza v
16	132	54.3	566	12 Q82496	Q82496 Influenzavi

17	132	54.3	571	12 Q03909	Q03909 Influenza a
18	131	53.9	109	12 Q67050	Q67050 Influenzavi
19	131	53.9	109	12 Q67053	Q67053 Influenzavi
20	131	53.9	109	12 Q67051	Q67051 Influenzavi
21	131	53.9	109	12 Q67052	Q67052 Influenzavi
22	131	53.9	362	12 Q9QKD3	Q9QKD3 Influenza a
23	131	53.9	362	12 Q9QKD1	Q9QKD1 Influenza a
24	131	53.9	362	12 Q82513	Q82513 Influenzavi
25	131	53.9	362	12 Q9OKD2	Q9OKD2 Influenza a
26	131	53.9	362	12 Q84174	Q84174 Influenzavi
27	131	53.9	362	12 Q82517	Q82517 Influenzavi
28	131	53.9	365	12 Q9DL25	Q9DL25 Influenza a
29	131	53.9	367	12 Q9DL22	Q9DL22 Influenza a
30	131	53.9	368	12 Q9DL29	Q9DL29 Influenza a
31	131	53.9	369	12 Q9DL26	Q9DL26 Influenza a
32	131	53.9	369	12 P87689	P87689 Influenzavi
33	131	53.9	369	12 Q9DL06	Q9DL06 Influenza a
34	131	53.9	371	12 Q9DL24	Q9DL24 Influenza a
35	131	53.9	371	12 P87685	P87685 Influenzavi
36	131	53.9	373	12 Q9DL20	Q9DL20 Influenza a
37	131	53.9	374	12 Q9DL21	Q9DL21 Influenza a
38	131	53.9	375	12 Q9DL27	Q9DL27 Influenza a
39	131	53.9	375	12 Q9DL05	Q9DL05 Influenza a
40	131	53.9	376	12 Q9DL30	Q9DL30 Influenza a
41	131	53.9	376	12 Q9DL04	Q9DL04 Influenza a
42	131	53.9	377	12 Q9E7P6	Q9E7P6 Influenza a
43	131	53.9	382	12 Q9DL03	Q9DL03 Influenza a
44	131	53.9	408	12 Q9E7P5	Q9E7P5 Influenza a
45	131	53.9	409	12 Q9Q0L5	Q9Q0L5 Influenza a

ALIGNMENTS

RESULT 1

Q8JRK63 PRELIMINARY; PRT; 384 AA.

AC Q8JRK63;

DT 01-OCT-2002 (TREMBLrel. 22. Created)

DT 01-OCT-2002 (TREMBLrel. 22. Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23. Last annotation update)

DE Hemagglutinin (Fragment).

GN H3HA.

OS Influenza A virus (A/teal/Germany/W201r/01).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

OX NCBI_TaxID=205472;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/teal/Germany/W201r/01;

RA Werner O., Starick E., Mueller T., Muehle R.;

RT "Characterisation of avian influenza virus isolates from wild birds from Germany."

RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC EMBL; AJ506781; CAD44999.1; -.

DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTIN12.

DR PRODOM; PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

FT NON TER 384

FT SEQUENCE 384 AA; 42076 MW; 459731795CASCCE38 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 384;

Best Local Similarity 100.0%; Pred. No. 9.5e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCLFGALAGFIENGWEGMIDGWY 24

DB 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 2

Q98052 PRELIMINARY; PRT; 566 AA.
AC 098052; 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hemagglutinin precursor (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OK NCB1_taxid=197911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81053698; PubMed=6253883;
RA Both G.W., Sleight M.J.;
RT "Complete nucleotide sequence of the haemagglutinin gene from a human influenza virus of the Hong Kong subtype.";
RL Nucleic Acids Res. 8:2561-2575(1980).
RN [2]
RP SEQUENCE OF 17-344 FROM N.A.
RX MEDLINE=81194918; PubMed=6164798;
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza subtype: Correlation of amino acid changes with alterations in viral antigenicity.";
RL J. Virol. 37:845-853(1981).
RN [3]
RP SEQUENCE OF 17-566 FROM N.A.
RX MEDLINE=82033276; PubMed=6169843;
RA Both G.W., Sleight M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift.";
RL J. Virol. 39:845-853(1981).
RN [4]
RP FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DE EMBL: J02135; AAA43189.1; -.
DR HSP; P03437; 1HGE.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CHAIN 346 566
FT CHAIN 566 63414 MW; C447FDP65B84FCF9 CRC64;
SQ SEQUENCE 566 AA; 63414 MW; C447FDP65B84FCF9 CRC64;
Query Match 55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

Q8U251 PRELIMINARY; PRT; 566 AA.
AC 08U251; 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Haemagglutinin.
GN HA.

OS Influenza A virus (A/pet bird/Hong Kong/1559/99 (H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OK NCB1_taxid=183665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/pet bird/Hong Kong/1559/99;
RA Chin P., Shortridge K.F.;
RT "Characterisation of avian H3 influenza viruses.";
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DE EMBL; AJ427304; CAD20336.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63403 MW; F11C91B6A0183484 CRC64;

Query Match

55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24

DB 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 4

Q8QLN8 PRELIMINARY; PRT; 566 AA.
AC Q8QLN8; 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Haemagglutinin.
GN HA.
OS Influenza A virus (A/aquatic bird/Hong Kong/399/99 (H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OK NCB1_taxid=183664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/aquatic bird/Hong Kong/399/99;
RA Chin P., Shortridge K.F.;
RT "Characterisation of influenza viruses from wild aquatic birds.";
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DE EMBL; AJ427297; CAD20332.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63412 MW; 68913C222C97B92E CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24

DB 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 5

Q67132
ID 067132 PRELIMINARY; PRT; 566 AA.
AC 067132;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=150147;
RN [1]
RC STRAIN=A/Aichi/2/68;
RA Min J.W., Verhoeven M., Fang R.-X., Devos R., Huylebroeck D.,
RA Fiers W.,
RT "Shift and drift in influenza viruses."
RL (In) Carlile M.J., Collins J.F., Moseley B.E. B. (eds.);
RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,
RL Cambridge University Press, New York (1981)
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: M55059; AAA43239.1; -.
DR HSSP; P03437; IHGE.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin.1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT CHAIN 1 344 HEMAGGLUTININ.
FT CHAIN 2 566 NEURAMINIDASE.
SQ SEQUENCE 566 AA; 63441 MW; ESD1B97DP96FECA CRC64;
Query Match 55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIGFIENGWEGMIDGMYG 24
Db 346 GLFGAIGFIENGWEGMIDGMYG 368
RESULT 6
ID 067125 PRELIMINARY; PRT; 566 AA.
AC 067125;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Seal/Ma/3911/92;
RX MEDLINE=95146951; PubMed=7844533;
RA Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT "The appearance of H3 influenza viruses in seals."
J. Gen. Virol. 76:199-203 (1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; L31949; AAA64229.1; -.
DR HSSP; P03437; 2V1U.

DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin.1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63456 MW; AE556302A9EBB99F CRC64;
Query Match 55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIGFIENGWEGMIDGMYG 24
Db 346 GLFGAIGFIENGWEGMIDGMYG 368
RESULT 7
ID 08UXR3 PRELIMINARY; PRT; 566 AA.
AC 08UXR3;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (A/swine/Potsdam/35/82 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=183769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/swine/Potsdam/35/82;
RA Groetinger I., Sues U., Groetinger C.;
RT "Evolution of european human and porcine influenza viruses."
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ252132; CAC81018.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin.1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63529 MW; 6AAA4C84B4DDE68A CRC64;
Query Match 55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIGFIENGWEGMIDGMYG 24
Db 346 GLFGAIGFIENGWEGMIDGMYG 368
RESULT 8
ID 091MA7 PRELIMINARY; PRT; 566 AA.
AC 091MA7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Hemagglutinin.
OS Influenza A virus (A/Hong Kong/1/68 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=108859;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Hong Kong/1/68 (H3N2);
RX MEDLINE=21287244; PubMed=11371620;

RA Brown E.G., Liu H., Kit L.C., Baird S., Neerallah M.;
 RT "Pattern of mutation in the genome of influenza A virus on adaptation
 RT to increased virulence in the mouse lung: identification of functional
 RT themes";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF348176; AAKS1718.1; -.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR Prodom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63356 MW; 01BBD0465BE158E1 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GLFGAAGFIENGWEGMIDGWY 24
 Db 346 GLFGAAGFIENGWEGMIDGWY 368

RESULT 9
 ID Q9DHG0 PRELIMINARY; PRT; 566 AA.
 AC Q9DHG0;
 DT 01-MAR-2001 (TRENBLREL. 16, Created)
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
 DE Haemagglutinin precursor.
 OS Influenza A virus H3N2.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CX NCBI_TaxID=41857;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=clone 7a;
 RA Mohsin M.A., Morris S.J., Smith H., Sweet C.;
 RT "Influenza virus-induced apoptosis: a dual role for viral
 RT neuraminidase";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AJ289703; CAC18525.1; -.
 DR HSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR Prodom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
 FT SIGNAL 16 POTENTIAL.
 SQ SEQUENCE 566 AA; 63356 MW; 0BA681929300F72F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GLFGAAGFIENGWEGMIDGWY 24
 Db 346 GLFGAAGFIENGWEGMIDGWY 368

RESULT 10
 ID Q910M5 PRELIMINARY; PRT; 566 AA.

AC Q910M5;
 DT 01-DEC-2001 (TRENBLREL. 19, Created)
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Hemagglutinin.
 OS Influenza A virus (A/Hong Kong/1/68 (H3N2)).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 CX NCBI_TaxID=108859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Hong Kong/1/68 (H3N2);
 RX MEDLINE=21287244; PubMed=11371620;
 RA Brown E.G., Liu H., Kit L.C., Baird S., Neerallah M.;
 RT "Pattern of mutation in the genome of influenza A virus on adaptation
 RT to increased virulence in the mouse lung: identification of functional
 RT themes";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF348179; AAKS1721.1; -.
 DR EMBL; AF348177; AAKS1719.1; -.
 DR EMBL; AF348178; AAKS1720.1; -.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR Prodom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF1EEB9F4 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GLFGAAGFIENGWEGMIDGWY 24
 Db 346 GLFGAAGFIENGWEGMIDGWY 368

RESULT 11
 ID Q67126 PRELIMINARY; PRT; 566 AA.
 AC Q67126;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
 DE Hemagglutinin.
 GN HA.
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses.
 CX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Seal/Ma/3964/92;
 RX MEDLINE=95146951; PubMed=7844533;
 RA Callan R.J., Early G., Kida H., Hinshaw V.S.;
 RT "The appearance of H3 influenza viruses in seals";
 RL J. Gen. Virol. 76:199-203(1995).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; L32024; AAB64228.1; -.
 DR HSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR Prodom; PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CEB7D08 CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
 |||||
 DB 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 12
 O9DXE3 PRELIMINARY; PRT; 301 AA.
 AC O9DXE3; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin (Fragment).
 GN HA1.
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=140665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Shorebird/Taiwan/31-4/99;
 RA Lee M.S., Cheng P.C., Shieh J.H., Cheng M.C., Lee L.H., Shieh H.K.;
 RT "Identification and subtyping of avian influenza virus by reverse
 transcription-polymerase chain reaction.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF111750; AAG3016.1; -.
 DR InterPro; IPR001364; Hemagglutn;
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; HEMAGGLUTN; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT NON_TER 301
 SQ SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;
 Query Match 54.7%; Score 133; DB 12; Length 301;
 Best Local Similarity 95.7%; Pred. No. 9.7e-08;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
 |||||
 DB 250 GLFGAIAAGFIENGWEGMIDGWYG 272

RESULT 13
 O82499 PRELIMINARY; PRT; 550 AA.
 AC O82499; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin HA1 and HA2 (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82/BS;
 RA Hartley C.A., Ward A.C., Anders E.M.;
 RT "Virulence of influenza virus for mice is associated with loss of

RT oligosaccharide from the hemagglutinin molecule.";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; U08859; AAA18782.1; -.
 DR HSSP; P03437; 2YIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT CHAIN 1
 FT CHAIN 328
 FT CHAIN 550
 SQ SEQUENCE 550 AA; 61772 MW; 50BD62B6BFE11PD8 CRC64;
 Query Match 54.3%; Score 132; DB 12; Length 550;
 Best Local Similarity 95.7%; Pred. No. 2.3e-07;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
 |||||
 DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 14
 O82498 PRELIMINARY; PRT; 550 AA.
 AC O82498; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin HA1 and HA2 (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82;
 RA Hartley C.A., Ward A.C., Anders E.M.;
 RT "Virulence of influenza virus for mice is associated with loss of
 RT oligosaccharide from the hemagglutinin molecule.";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; U08858; AAA18781.1; -.
 DR HSSP; P03437; 2YIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT CHAIN 1
 FT CHAIN 328
 FT CHAIN 550
 SQ SEQUENCE 550 AA; 61802 MW; 114413B1CE5A1F6A CRC64;
 Query Match 54.3%; Score 132; DB 12; Length 550;

Best Local Similarity 95.7%; Pred. No. 2.3e-07;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAIAGFIENGEGMIDGWYG 24
|:|||||
Db 330 GFGAIAGFIENGEGMIDGWYG 352

RESULT 15

Q82753 PRELIMINARY; PRT; 550 AA.
AC Q82753;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Haemagglutinin (Fragment).
OS Influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Unclassified Orthomyxoviridae.
OX NCBI_TaxID=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82/BS/ML10;
RX MEDLINE=97300854; PubMed=9155874;
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)
RT virus associated with increased virulence for mice.";
RL Arch. Virol. 142:75-88 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82/BS/ML10;
RX MEDLINE=97456249; PubMed=9311563;
RA Ward A.C.;
RT "Virulence of Influenza A virus for mouse lung.";
RL Virus Genes 14:187-194 (1997).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; U08905; AAC79579.1; -.
DR HSSP; P03437; 2VIT.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 1
FT CHAIN 1 328 HAEMAGGLUTININ HA1.
FT CHAIN 330 550 HAEMAGGLUTININ HA2.
SQ SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

Query Match 54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 2.3e-07;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAIAGFIENGEGMIDGWYG 24
|:|||||
Db 330 GFGAIAGFIENGEGMIDGWYG 352

Search completed: January 30, 2004, 00:24:41
Job time : 45.446 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:28 ; Search time 21 Seconds
(without alignments)
36.636 Million cell updates/sec

Title: SEQ10
Perfect score: 38
Sequence: 1 sinifek1 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	386	1 OACH	ovalbumin [validated]
2	34	89.5	627	2 S37954	RNA polymerase I t
3	33	86.8	168	3 F90095	hypothetical prote
4	33	86.8	315	2 T10818	1-aminocyclopropan
5	33	86.8	383	2 S11433	ovalbumin - Japanese
6	32	84.2	260	2 G87349	conserved hypothet
7	32	84.2	347	2 A99989	cyclin B (imported
8	32	84.2	1305	2 T23314	hypothetical prote
9	31	81.6	244	2 T06961	ABC transpore prot
10	31	81.6	304	2 C64109	site-specific DNA-
11	31	81.6	307	2 E71206	hypothetical prote
12	31	81.6	309	2 H90548	hypothetical prote
13	31	81.6	364	2 A84726	probable poly(ADP-
14	31	81.6	397	2 T34441	hypothetical prote
15	31	81.6	680	2 T42923	infected cell prot
16	30	78.9	142	2 A83093	50S ribosomal prot
17	30	78.9	232	1 DXCH	ovalbumin-related
18	30	78.9	347	2 A82358	hypothetical prote
19	30	78.9	349	2 H71923	probable inner mem
20	30	78.9	378	2 T18486	hypothetical prote
21	30	78.9	440	2 F70117	ATP-dependent prote
22	30	78.9	467	2 AF1317	hypothetical prote
23	30	78.9	470	2 D84614	conserved hypothet
24	30	78.9	481	2 B70179	hypothetical prote
25	30	78.9	487	2 C97144	probable membrane
26	30	78.9	520	2 B82206	probable purine-bi
27	30	78.9	558	2 S61604	hypothetical prote
28	30	78.9	610	2 T25262	hypothetical prote
29	30	78.9	842	2 E72373	hypothetical prote

30	30	78.9	845	2 D90130	hypothetical prote
31	30	78.9	1004	2 B59483	hypothetical prote
32	30	78.9	1115	2 B84476	probable TPR repea
33	30	78.9	1163	2 G97236	ATPase involved in
34	30	78.9	2116	2 T49818	glutamate synthase
35	30	78.9	2469	2 H36812	hypothetical prote
36	30	78.9	4092	1 S38128	dynein heavy chain
37	29	76.3	69	2 A96696	protein PIN2.2 (1
38	29	76.3	88	4 S54022	hypothetical prote
39	29	76.3	110	2 F97094	probable transcrip
40	29	76.3	141	2 H69044	peptidylprolyl iso
41	29	76.3	143	2 S45537	peptidylprolyl iso
42	29	76.3	170	2 G68356	hypothetical prote
43	29	76.3	209	2 D86882	hypothetical prote
44	29	76.3	237	2 F64506	2-hydroxyhepta-2,4
45	29	76.3	275	1 C69030	Mu1225 protein hom

ALIGNMENTS

RESULT 1

OACH
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1979 #sequence revision 30-Jun-1993 #text change 15-Sep-2000
C:Accession: A90455; I50402; I50605; A93197; A93827; A90092; A90092; A61297; A42793; A01
R:Woo, S.L.C.; Beattie, W.G.; Catterall, J.F.; Dugalczyk, A.; Staden, R.; Brownlee, G.G
Biochemistry 20, 6437-6446, 1981
A>Title: Complete nucleotide sequence of the chicken chromosomal ovalbumin gene and its
A:Reference number: A90455; MUID:82069038; PMID:6272839
A:Accession: A90455
A:Molecule type: DNA
A:Residues: 1-386 <MO>
A:Cross-References: EMBL:V00438; NID:G63719; PIDN:CAA23716.1; PID:G808974
A:Note: A number of silent polymorphic sites are identified and discussed
A:Note: Thr-188 is also predicted
R:Catterall, J.F.; O'Malley, B.W.; Robertson, M.A.; Staden, R.; Tanaka, Y.; Brownlee, G
Nature 275, 510-513, 1978
A>Title: Nucleotide sequence homology at 12 intron-exon junctions in the chick ovalbumin
A:Reference number: I50402; MUID:79010682; PMID:692731
A:Accession: I50402
A:Molecule type: DNA
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-386 <CAT>
A:Cross-References: GB:M34352; NID:G212501; PIDN:AAA48998.1; PID:G212503
R:Robertson, M.A.; Staden, R.; Tanaka, Y.; Catterall, J.F.; Brownlee, G.
Nature 278, 370-372, 1979
A>Title: Sequence of three introns in the chick ovalbumin gene.
A:Reference number: I50605; MUID:79135070; PMID:423993
A:Accession: I50605
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4, 'A', '6-118, 'F', '120-155 <ROB>
A:Cross-References: EMBL:V00382; NID:G63051; PIDN:CAA23681.1; PID:G63052
R:McReynolds, L.; O'Malley, B.W.; Nisbet, A.D.; Fothergill, J.E.; Givoli, D.; Fields, S.,
Nature 273, 723-728, 1978
A>Title: Sequence of chicken ovalbumin mRNA.
A:Reference number: A93197; MUID:78199842; PMID:661981
A:Accession: A93197
A:Molecule type: mRNA
A:Residues: 1-386 <MCB>
A:Cross-References: EMBL:V00383; NID:G63053
A:Note: a minor component has Asp-312
R:Palmiter, R.D.; Gagnon, J.; Walsh, K.A.
Proc. Natl. Acad. Sci. U.S.A. 75, 94-98, 1978
A>Title: Ovalbumin: a secreted protein without a transient hydrophobic leader sequence.
A:Reference number: A93827; MUID:78116057; PMID:272676
A:Accession: A93827
A:Molecule type: Protein
A:Residues: 2-33, 'X', '35-36 <PAL>
R:Thompson, E.O.P.; Fisher, W.K.
Aust. J. Biol. Sci. 31, 443-446, 1978

A>Title: A correction and extension of the acetylated amino terminal sequence of ovalbumin
 A:Reference number: A90093; MUID:79186958; PMID:751625
 A:Accession: A90093
 A:Molecule type: protein
 A:Residues: 2-17 <TH1>
 R:Thompson, E.O.P.; Fisher, W.K.
 Aust. J. Biol. Sci. 31, 433-442, 1978
 A>Title: Amino acid sequences containing half-cysteine residues in ovalbumin.
 A:Reference number: A90092; MUID:79186957; PMID:751624
 A:Accession: A90092
 A:Molecule type: protein
 A:Residues: 6-17,30-36,61-79,116-124,367-374,380-386 <TH2>
 R:Tsunasawa, S.; Narita, K.
 J. Biochem. 92, 607-613, 1982
 A>Title: Micro-identification of amino-terminal acetylamino acids in proteins.
 A:Reference number: A61297; MUID:83056735; PMID:6754709
 A:Accession: A61297
 A:Molecule type: protein
 A:Residues: 2-6 <TSU>
 R:Takahashi, N.; Hirose, M.
 J. Biol. Chem. 267, 11565-11572, 1992
 A>Title: Reversible denaturation of disulfide-reduced ovalbumin and its reoxidation gene
 A:Reference number: A42793; MUID:92283876; PMID:1597484
 A:Accession: A42793
 A:Molecule type: protein
 A:Residues: 60-73, 'X', 75-85,112-119, 'EX', 122-123 <TKA>
 R:Stein, P.E.; Leslie, A.G.W.
 submitted to the Brookhaven Protein Data Bank, November 1990
 A:Reference number: A50294; PDB:1OVA
 A:Contents: annotation; X-ray crystallography; 1.95 angstroms, residues 2-386
 R:Stein, P.E.; Leslie, A.G.W.; Finch, J.T.; Carrell, R.W.
 J. Mol. Biol. 221, 941-959, 1991
 A>Title: Crystal structure of unbleached ovalbumin at 1.95 Angstroms resolution.
 A:Reference number: A58761; MUID:92046044; PMID:1942038
 A:Contents: annotation; X-ray crystallography; 1.95 angstroms
 C:Genetics:
 A:Introns: 56/3; 73/3; 116/3; 156/1; 203/3; 255/3
 C:Superfamily: antithrombin III
 C:Keywords: acetylated amino end; glycoprotein; phosphoprotein
 F:2-184/Product: ovalbumin #status experimental <MAT>
 F:2/Modified site: acetylated amino end (Gly) (in mature form) #status experimental
 F:69,345/Binding site: phosphate (Ser) (covalent) #status experimental
 F:74-121/Disulfide bonds: #status experimental
 F:293/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 100.0%; Score 38; DB 1; Length 386;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 Db 258 SIINFEXL 265

RESULT 2
 S37954
 RNA polymerase I transcription factor RRN3 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YKL125W
 C:Species: Saccharomyces cerevisiae
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 13-Mar-1998
 C:Accession: S37954; S71600
 R:Ramazani Rad, M.; Xu, G.; Kirchrath, L.; Fritze, C.; Keuchel, H.; Hollenberg, C.P.
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S37953
 A:Accession: S37954
 A:Molecule type: DNA
 A:Residues: 1-627 <RAM>
 A:Cross-references: EMBL:Z28125; NID:9486210; PID:9486211; MIPS:YKL125W
 A:Experimental source: strain S288C
 R:Yamamoto, R.T.; Nogi, Y.; Dodd, J.A.; Nomura, M.
 EMBO J. 15, 3964-3973, 1996
 A>Title: RRN3 gene of Saccharomyces cerevisiae encodes an essential RNA polymerase I tra
 A:Reference number: S71600; MUID:96324404; PMID:8670901

A:Accession: S71600
 A:Molecule type: DNA
 A:Residues: 1-627 <YAM>
 C:Genetics:
 A:Gene: SGD:RRN3
 A:Cross-references: SGD:S0001608; MIPS:YKL125W
 A:Map position: 11L
 C:Keywords: nucleus

Query Match 89.5%; Score 34; DB 2; Length 627;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 7
 |||||
 Db 118 SIINFEXL 124

RESULT 3
 P90095
 hypothetical protein orf168 [imported] - Guillardia theta nucleomorph
 C:Species: nucleomorph Guillardia theta
 A>Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: P90095
 R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re
 Nature 410, 1091-1096, 2001
 A>Title: The highly reduced genome of an enslaved algal nucleus.
 A:Reference number: A99082; MUID:11323671; PMID:11323671
 A:Accession: P90095
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-168 <DOU>
 A:Cross-references: GB:AF165818; NID:913794518; PIDN:AAK39893.1; GSPDB:GN00150
 C:Genetics:
 A:Gene: orf168
 A:Map position: 1
 A:Genome: nucleomorph
 C:Keywords: nucleomorph

Query Match 86.8%; Score 33; DB 2; Length 168;
 Best Local Similarity 75.0%; Pred. No. 11;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 :|||:
 Db 108 SIINFEXL 115

RESULT 4
 T10818
 1-aminocyclopropane-1-carboxylate oxidase (EC 1.4.3.-) - kidney bean
 C:Species: Phaseolus vulgaris (kidney bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 19-May-2000
 C:Accession: T10818
 R:Pidgeon, C.M.; Facchini, P.J.; Reid, D.M.
 submitted to the EMBL Data Library, March 1998
 A:Reference number: Z17172
 A:Accession: T10818
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-315 <PID>
 A:Cross-references: EMBL:AF053354; NID:93037046; PIDN:AACT2934.1; PID:93037047
 A:Experimental source: cultivar Taylor bush bean, leaf
 C:Genetics:
 A:Gene: ACO1
 C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase
 C:Keywords: ethylene biosynthesis; iron; metalloprotein; oxidoreductase
 F:39,177,23/Binding site: iron (His) #status predicted

Query Match 86.8%; Score 33; DB 2; Length 315;
 Best Local Similarity 85.7%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEKL 8
:|||||
Db 6 VINFEKL 12

RESULT 5

S11433
ovalbumin - Japanese quail
C/Species: Coturnix coturnix japonica (Japanese quail)
C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C/Accession: S11433
R/Mucha, J.; Kaudiny, J.; Kladinyova, V.; Hanes, J.; Simuth, J.
Nucleic Acids Res. 18, 5553, 1990
A/Title: The sequence of Japanese quail ovalbumin cDNA.
A/Reference number: S11433; MUID:91016850; PMID:2216734
A/Accession: S11433
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-383 <MUC>
A/Cross-references: EMBL:X53964; NID:g62643; PIDN:CAA37916.1; PID:g62644
C/Superfamily: antithrombin III

Query Match 86.8%; Score 33; DB 2; Length 383;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEKL 8
:|||||
Db 258 STINFEKL 265

RESULT 6

G87349
conserved hypothetical protein CC0810 [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: G87349
R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolar
n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: G87349
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-260 <STO>
A/Cross-references: GB:AE005673; NID:g13422057; PIDN:AAK2795.1; GSPDB:GN00148
C/Genetics:

Query Match 84.2%; Score 32; DB 2; Length 260;
Best Local Similarity 87.5%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STINFEKL 8
:|||||
Db 215 STINFEKL 222

RESULT 7

A99989
cyclin B [imported] - Guillardia theta nucleomorph
C/Species: nucleomorph guillardia theta
A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C/Accession: A99989
R/Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re
Nature 410, 1091-1096, 2001
A/Title: The highly reduced genome of an enslaved algal nucleus.
A/Reference number: A99082; MUID:11323671; PMID:11323671
A/Accession: A99989

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-347 <DOU>
A/Cross-references: GB:AF165818; NID:g13794469; PIDN:AAK39844.1; GSPDB:GN00150
C/Genetics:
A/Gene: cycB
A/Map position: 1
A/Genome: nucleomorph
C/Keywords: nucleomorph

Query Match 84.2%; Score 32; DB 2; Length 347;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEKL 8
:|||||
Db 86 NVINFEKL 93

RESULT 8

T23314
hypothetical protein T14G10.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C/Accession: T23314; T24919
R/Wild, A.
submitted to the EMBL Data Library, February 1996
A/Reference number: Z19725
A/Accession: T23314
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1305 <WIL>
A/Cross-references: EMBL:Z69664; PIDN:CAA3519.1; GSPDB:GN00022; CESP:T14G10.2
A/Experimental source: clone K04D7
R/Wild, A.
submitted to the EMBL Data Library, January 1996
A/Reference number: Z19954
A/Accession: T24919
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1305 <WIL>
A/Cross-references: EMBL:Z68880; PIDN:CAA33100.1; GSPDB:GN00022; CESP:T14G10.2
A/Experimental source: clone T14G10
C/Genetics:
A/Gene: CESP:T14G10.2
A/Map position: 4
A/Introns: 450/1; 463/2; 696/2; 763/2; 843/2; 935/3; 1012/1; 1091/1; 1143/1; 1189/2; 125

Query Match 84.2%; Score 32; DB 2; Length 1305;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEKL 8
:|||||
Db 903 IINFEKL 909

RESULT 9

T06961
ABC transport protein homolog - Cyanophora paradoxa cyanelle
C/Species: cyanelle Cyanophora paradoxa
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 17-Mar-2000
C/Accession: T06961
R/Stewart, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohner, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A/Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A/Reference number: Z15840
A/Accession: T06961
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-244 <STI>
A/Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81304.1; PID:g1016217
A/Experimental source: strain Pfingsheim LB555

C:Genetics:
 A:Genome: cyanelle
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: cyanelle

Query Match 81.6%; Score 31; DB 2; Length 244;
 Best Local Similarity 75.0%; Pred. No. 44;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEK 8
 :|||||:
 Db 204 STINDEL 211

RESULT 10
 C64109
 site-specific DNA-methyltransferase (cytosine-specific) (EC 2.1.1.73) - Haemophilus inf
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000
 C:Accession: C64109
 R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: C64109
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-304 <TIGR>
 A:Cross-references: GB:142023; NID:93212210; PIDN:ANC22700.1; PID:g574073; T
 C:Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII
 C:Keywords: methyltransferase; restriction modification system; S-adenosylmethionine

Query Match 81.6%; Score 31; DB 2; Length 304;
 Best Local Similarity 85.7%; Pred. No. 56;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEK 7
 :|||||:
 Db 291 ATINFEK 297

RESULT 11
 E71206
 hypothetical protein PH1919 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
 C:Accession: E71206
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Seki,
 N.; Ohnuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuhida, N.; Oguchi,
 DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
 A:Reference number: A71000; MUID:98344137; PMID:9679194
 A:Accession: E71206
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-307 <KAW>
 A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BA031044.1; PID:g3258361
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH1919

Query Match 81.6%; Score 31; DB 2; Length 307;
 Best Local Similarity 75.0%; Pred. No. 57;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEK 8
 :|||||:
 Db 3 NINFEEL 10

RESULT 12
 H90548
 hypothetical protein MYPu_2960 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: H90548
 R:Chabaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: H90548
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <KUR>
 A:Cross-references: GB:AL445566; PID:g14089710; PIDN:CA013469.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYPu_2960
 A:Genetic code: SCC3

Query Match 81.6%; Score 31; DB 2; Length 309;
 Best Local Similarity 71.4%; Pred. No. 57;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEK 7
 :|||||:
 Db 287 STINFEK 293

RESULT 13
 A64726
 probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: A64726
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKien, S.E.; Umayam, L.; Tallon, L.
 eus, D.; Nierman, W.C.; White, O.; Eilen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A64420; MUID:20083487; PMID:10617197
 A:Accession: A64726
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-364 <STO>
 A:Cross-references: GB:AE002093; NID:g4887750; PIDN:AA032286.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g31860
 A:Map position: 2

Query Match 81.6%; Score 31; DB 2; Length 364;
 Best Local Similarity 85.7%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEK 8
 :|||||:
 Db 88 IINFEK 94

RESULT 14
 T34441
 hypothetical protein K11H12.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34441
 R:Bradshaw, H.
 submitted to the EMBL Data Library, February 1997
 A:Description: The sequence of C. elegans coenid K11H12.
 A:Reference number: Z21526
 A:Accession: T34441
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-397 <BRA>
 A:Cross-references: EMBL:U88168; PIDN:AAC24400.1; GSPDB:GN00022; CESP:K11H12.3
 A:Experimental source: strain Bristol N2; clone K11H12
 C:Genetics:
 A:Gene: CESP:K11H12.3
 A:Map position: 4
 A:introns: 39/3; 68/2; 118/1; 206/2; 280/3

Query Match 81.6%; Score 31; DB 2; Length 397;
 Best Local Similarity 75.0%; Pred. No. 75;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 |||||
 Db 190 SIINFESKI 197

RESULT 15

T42923
 infected cell protein - ateline herpesvirus 3 (strain 73)
 C:Species: ateline herpesvirus 3
 A:Variety: strain 73
 C:Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text_change 05-May-2000
 C:Accession: T42923
 R:Albrecht, J.C.; Fleckenstein, B.
 submitted to the EMBL Data Library, August 1998
 A:Description: Primary structure of the herpesvirus ateles genome.
 A:Reference number: Z22274
 A:Accession: T42923
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-680 <ALB>
 A:Cross-references: EMBL:AF083424; PIDN:AAC95539.1
 A:Experimental source: strain 73
 C:Genetics:
 A:Note: orf07
 C:Superfamily: herpesvirus infected cell protein ICP18.5

Query Match 81.6%; Score 31; DB 2; Length 680;
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEX 7
 |::|||
 Db 618 SVLNFEK 624

Search completed: January 30, 2004, 07:09:46
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compgen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:27 ; Search time 12 Seconds
(without alignments)
31.351 Million cell updates/sec

Title: SEQ10
Perfect score: 38
Sequence: 1 sinife1 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	385	1 OVAL_CHICK	P01012 gallus galli
2	34	89.5	627	1 RNN3_YEAST	P36070 saccharomyc
3	33	86.8	382	1 OVAL_COTJA	P19104 coturnix co
4	32	84.2	267	1 RMO9_HUMAN	O9byd2 homo sapien
5	31	81.6	242	1 TPIS_MYCEL	P48779 mycoplasma
6	31	81.6	244	1 YCSD_CYARA	P48334 mycoplasma
7	31	81.6	304	1 MTH5_HAEIN	P45000 haemophilus
8	30	78.9	231	1 RLI_BUCAP	O8K67 buchnera ap
9	30	78.9	232	1 OVAX_CHICK	P01013 gallus galli
10	30	78.9	349	1 Y567_HELPJ	O921r4 heilobacte
11	30	78.9	364	1 SERC_DROME	O9van0 drosophila
12	30	78.9	527	1 PRCK_FUSNN	O8re12 fusobacteri
13	30	78.9	558	1 MNT2_YEAST	P53059 saccharomyc
14	30	78.9	842	1 DPA3_THEMA	O9zhg4 thermotoga
15	30	78.9	1163	1 SBCC_CLOAB	O97EK1 clostridium
16	30	78.9	2469	1 TEGU_HSVSA	O01056 herpesvirus
17	30	78.9	4092	1 DVHC_YEAST	P34622 saccharomyc
18	29	76.3	143	1 PRIB_BACSU	P35137 bacillus su
19	29	76.3	237	1 YG56_METUA	O59050 methanococc
20	29	76.3	394	1 CC91_YEAST	P41733 saccharomyc
21	29	76.3	445	1 RPN5_YEAST	O12250 saccharomyc
22	29	76.3	453	1 EX7L_RICPR	O9zcp8 rickettsia
23	29	76.3	482	1 V232_ROMPY	O9j503 fowlpox vir
24	29	76.3	489	1 T283_STPAU	P16667 staphylococ
25	29	76.3	749	1 PCRA_LEUCI	O98qg0 leuconostoc
26	29	76.3	1002	1 HPS3_MOUSE	O9jvb4 mus musculu
27	29	76.3	1233	1 SMCI_SCHPO	O94383 schizosacch
28	29	76.3	1234	1 YNKS_CABEL	P34578 caenorhabdi
29	29	76.3	2054	1 YCR2_PINTA	P41653 pinus thunb
30	28	73.7	102	1 CTT1_ORISA	P09229 oryza sativ
31	28	73.7	124	1 PA25_AGRKA	O42189 aglystodon
32	28	73.7	177	1 YASO_METUA	O58450 methanococc
33	28	73.7	214	1 PYRE_PASMU	O9cjw4 pasteurrella

34 28 73.7 222 1 BID2_YERPE O9agd4 yersinia pe
35 28 73.7 240 1 YDED_SCHPO O10446 schizosacch
36 28 73.7 242 1 TPIS_MYCH P50920 mycoplasma
37 28 73.7 269 1 HIS9_LACLA O02150 lactococcus
38 28 73.7 278 1 TNP6_RAT P35940 rattus norv
39 28 73.7 279 1 TNP6_MOUSE P41047 mus musculu
40 28 73.7 306 1 COA4_STRP3 O8K7c7 streptococc
41 28 73.7 306 1 COA4_STRP8 O8D0v9 streptococc
42 28 73.7 306 1 COA4_STRP9 O992h1 streptococc
43 28 73.7 309 1 LDH2_LACPL P59390 lactobacill
44 28 73.7 315 1 ACC1_LYCES P05116 lycopersico
45 28 73.7 316 1 ACC2_LYCES P07920 lycopersico

ALIGNMENTS

RESULT 1
ID OVAL_CHICK STANDARD; PRT; 385 AA.
AC P01012;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ovalbumin (Plakalbumin) (Allergen Gal d 2) (Gal d II).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OK NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82069038; PubMed=6272839;
RA Brownlee G.G., Beattie W.G., Catterall J.F., Dugaiczky A., Staden R.,
RA Woo S.L.C., Beattie W.G., O'Malley B.W.,
RT "Complete nucleotide sequence of the chicken chromosomal ovalbumin
RT gene and its biological significance.",
RL Biochemistry 20:6437-6446(1981).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=78199842; PubMed=661981;
RA McReynolds L., O'Malley B.W., Nisbet A.D., Fothergill J.E., Givol D.,
RA Fields S., Robertson M., Brownlee G.G.,
RT "Sequence of chicken ovalbumin mRNA.",
RL Nature 273:723-728(1978).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=79010682; PubMed=692731;
RA Catterall J.F., O'Malley B.W., Robertson M.A., Staden R.,
RA Tanaka Y., Brownlee G.G.,
RT "Nucleotide sequence homology at 12 intron-exon junctions in the
RT chick ovalbumin gene.",
RL Nature 275:510-513(1978).
RN [4]
RP SEQUENCE OF 1-35.
RX MEDLINE=78116057; PubMed=272676;
RA Palmiter R.D., Gagnon J., Walsh K.A.,
RT "Ovalbumin: a secreted protein without a transient hydrophobic leader
RT sequence.",
RL Proc. Natl. Acad. Sci. U.S.A. 75:94-98(1978).
RN [5]
RP SEQUENCE OF 1-16.
RX MEDLINE=79186958; PubMed=751625;
RA Thompson E.O.P., Fisher W.K.,
RT "A correction and extension of the acetylated amino terminal sequence
RT of ovalbumin.",
RL Aust. J. Biol. Sci. 31:443-446(1978).
RN [6]
RP SEQUENCE OF 5-16; 29-35; 60-78; 115-123; 366-373 AND 379-385.
RX MEDLINE=79186957; PubMed=751624;
RA Thompson E.O.P., Fisher W.K.,
RT "Amino acid sequences containing half-cysteine residues in ovalbumin.",
RL Aust. J. Biol. Sci. 31:433-442(1978).

```

RN [7]
RP N-TERMINUS ACETYLATION, AND PHOSPHORYLATION OF SER-68; SER-236 AND
RP SER-240.
RX MEDLINE=22056091; PubMed=12060738;
RA MacCoss M.J., McDonald W.H., Saraf A., Sadygov R., Clark J.M.,
RA Taato J.J., Gould K.L., Wolters D., Washburn M., Weiss A., Clark J.I.,
RA Yates J.R. III;
RT "Shotgun identification of protein modifications from protein
RT complexes and lens tissue."
RT Proc. Natl. Acad. Sci. U.S.A. 99:7900-7905(2002).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX MEDLINE=90370102; PubMed=2395463;
RA Stein P.E., Leslie A.G.W., Finch J.T., Turnell W.G., McLaughlin P.J.,
RA Carrell R.W.;
RT "Crystal structure of ovalbumin as a model for the reactive centre of
RT Serpin."
RT Nature 347:99-102(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX MEDLINE=92046044; PubMed=1942038;
RA Stein P.E., Leslie A.G.W., Finch J.T., Carrell R.W.;
RT "Crystal structure of uncleaved ovalbumin at 1.95-A resolution."
RT J. Mol. Biol. 221:941-959(1991).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=90278960; PubMed=2352279;
RA Wright H.T., Qian H.X., Huber R.;
RT "Crystal structure of plakalbumin, a proteolytically nicked form of
RT ovalbumin. Its relationship to the structure of cleaved alpha-1-
RT proteinase inhibitor."
RT J. Mol. Biol. 213:513-528(1990).
RN [11]
RP REVIEW.
RX MEDLINE=21312433; PubMed=11419711;
RA Huntington J.A., Stein P.E.;
RT "Structure and properties of ovalbumin."
RT J. Chromatogr. B 756:189-198(2001).
CC -1- FUNCTION: Not known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Major protein of egg white.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/O/OA.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00895; AAB59956.1; -
DR EMBL; V00438; CAA23716.1; -
DR EMBL; V00383; CAA23682.1; -
DR EMBL; M34352; AAA48998.1; -
DR EMBL; M34346; AAA48998.1; JOINED.
DR EMBL; M34347; AAA48998.1; JOINED.
DR EMBL; M34348; AAA48998.1; JOINED.
DR EMBL; M34349; AAA48998.1; JOINED.
DR EMBL; M34350; AAA48998.1; JOINED.
DR EMBL; M34351; AAA48998.1; JOINED.
DR F01; A90455; OACH.
DR PDB; 1OVA; 15-JUL-92.
DR PDB; 1JTI; 18-DEC-02.
DR PDB; 1VAC; 20-JUN-96.
DR GLYCOSULEDB; P01012; -
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; Serpin_1.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Serpin; Acetylation; Phosphorylation; Glycoprotein; 3D-structure;

```

```

KW Allergen.
FT INIT MET 0 0
FT MOD RES 1 1
FT MOD RES 68 68
FT MOD RES 73 73
FT MOD RES 76 76
FT MOD RES 77 77
FT MOD RES 78 78
FT MOD RES 79 79
FT MOD RES 80 80
FT MOD RES 81 81
FT MOD RES 82 82
FT MOD RES 83 83
FT MOD RES 84 84
FT MOD RES 85 85
FT MOD RES 86 86
FT MOD RES 87 87
FT MOD RES 88 88
FT MOD RES 89 89
FT MOD RES 90 90
FT MOD RES 91 91
FT MOD RES 92 92
FT MOD RES 93 93
FT MOD RES 94 94
FT MOD RES 95 95
FT MOD RES 96 96
FT MOD RES 97 97
FT MOD RES 98 98
FT MOD RES 99 99
FT MOD RES 100 100
FT MOD RES 101 101
FT MOD RES 102 102
FT MOD RES 103 103
FT MOD RES 104 104
FT MOD RES 105 105
FT MOD RES 106 106
FT MOD RES 107 107
FT MOD RES 108 108
FT MOD RES 109 109
FT MOD RES 110 110
FT MOD RES 111 111
FT MOD RES 112 112
FT MOD RES 113 113
FT MOD RES 114 114
FT MOD RES 115 115
FT MOD RES 116 116
FT MOD RES 117 117
FT MOD RES 118 118
FT MOD RES 119 119
FT MOD RES 120 120
FT MOD RES 121 121
FT MOD RES 122 122
FT MOD RES 123 123
FT MOD RES 124 124
FT MOD RES 125 125
FT MOD RES 126 126
FT MOD RES 127 127
FT MOD RES 128 128
FT MOD RES 129 129
FT MOD RES 130 130
FT MOD RES 131 131
FT MOD RES 132 132
FT MOD RES 133 133
FT MOD RES 134 134
FT MOD RES 135 135
FT MOD RES 136 136
FT MOD RES 137 137
FT MOD RES 138 138
FT MOD RES 139 139
FT MOD RES 140 140
FT MOD RES 141 141
FT MOD RES 142 142
FT MOD RES 143 143
FT MOD RES 144 144
FT MOD RES 145 145
FT MOD RES 146 146
FT MOD RES 147 147
FT MOD RES 148 148
FT MOD RES 149 149
FT MOD RES 150 150
FT MOD RES 151 151
FT MOD RES 152 152
FT MOD RES 153 153
FT MOD RES 154 154
FT MOD RES 155 155
FT MOD RES 156 156
FT MOD RES 157 157
FT MOD RES 158 158
FT MOD RES 159 159
FT MOD RES 160 160
FT MOD RES 161 161
FT MOD RES 162 162
FT MOD RES 163 163
FT MOD RES 164 164
FT MOD RES 165 165
FT MOD RES 166 166
FT MOD RES 167 167
FT MOD RES 168 168
FT MOD RES 169 169
FT MOD RES 170 170
FT MOD RES 171 171
FT MOD RES 172 172
FT MOD RES 173 173
FT MOD RES 174 174
FT MOD RES 175 175
FT MOD RES 176 176
FT MOD RES 177 177
FT MOD RES 178 178
FT MOD RES 179 179
FT MOD RES 180 180
FT MOD RES 181 181
FT MOD RES 182 182
FT MOD RES 183 183
FT MOD RES 184 184
FT MOD RES 185 185
FT MOD RES 186 186
FT MOD RES 187 187
FT MOD RES 188 188
FT MOD RES 189 189
FT MOD RES 190 190
FT MOD RES 191 191
FT MOD RES 192 192
FT MOD RES 193 193
FT MOD RES 194 194
FT MOD RES 195 195
FT MOD RES 196 196
FT MOD RES 197 197
FT MOD RES 198 198
FT MOD RES 199 199
FT MOD RES 200 200
FT MOD RES 201 201
FT MOD RES 202 202
FT MOD RES 203 203
FT MOD RES 204 204
FT MOD RES 205 205
FT MOD RES 206 206
FT MOD RES 207 207
FT MOD RES 208 208
FT MOD RES 209 209
FT MOD RES 210 210
FT MOD RES 211 211
FT MOD RES 212 212
FT MOD RES 213 213
FT MOD RES 214 214
FT MOD RES 215 215
FT MOD RES 216 216
FT MOD RES 217 217
FT MOD RES 218 218
FT MOD RES 219 219
FT MOD RES 220 220
FT MOD RES 221 221
FT MOD RES 222 222
FT MOD RES 223 223
FT MOD RES 224 224
FT MOD RES 225 225
FT MOD RES 226 226
FT MOD RES 227 227
FT MOD RES 228 228
FT MOD RES 229 229
FT MOD RES 230 230
FT MOD RES 231 231
FT MOD RES 232 232
FT MOD RES 233 233
FT MOD RES 234 234
FT MOD RES 235 235
FT MOD RES 236 236
FT MOD RES 237 237
FT MOD RES 238 238
FT MOD RES 239 239
FT MOD RES 240 240
FT MOD RES 241 241
FT MOD RES 242 242
FT MOD RES 243 243
FT MOD RES 244 244
FT MOD RES 245 245
FT MOD RES 246 246
FT MOD RES 247 247
FT MOD RES 248 248
FT MOD RES 249 249
FT MOD RES 250 250
FT MOD RES 251 251
FT MOD RES 252 252
FT MOD RES 253 253
FT MOD RES 254 254
FT MOD RES 255 255
FT MOD RES 256 256
FT MOD RES 257 257
FT MOD RES 258 258
FT MOD RES 259 259
FT MOD RES 260 260
FT MOD RES 261 261
FT MOD RES 262 262
FT MOD RES 263 263
FT MOD RES 264 264
FT MOD RES 265 265
FT MOD RES 266 266
FT MOD RES 267 267
FT MOD RES 268 268
FT MOD RES 269 269
FT MOD RES 270 270
FT MOD RES 271 271
FT MOD RES 272 272
FT MOD RES 273 273
FT MOD RES 274 274
FT MOD RES 275 275
FT MOD RES 276 276
FT MOD RES 277 277
FT MOD RES 278 278
FT MOD RES 279 279
FT MOD RES 280 280
FT MOD RES 281 281
FT MOD RES 282 282
FT MOD RES 283 283
FT MOD RES 284 284
FT MOD RES 285 285
FT MOD RES 286 286
FT MOD RES 287 287
FT MOD RES 288 288
FT MOD RES 289 289
FT MOD RES 290 290
FT MOD RES 291 291
FT MOD RES 292 292
FT MOD RES 293 293
FT MOD RES 294 294
FT MOD RES 295 295
FT MOD RES 296 296
FT MOD RES 297 297
FT MOD RES 298 298
FT MOD RES 299 299
FT MOD RES 300 300
FT MOD RES 301 301
FT MOD RES 302 302
FT MOD RES 303 303
FT MOD RES 304 304
FT MOD RES 305 305
FT MOD RES 306 306
FT MOD RES 307 307
FT MOD RES 308 308
FT MOD RES 309 309
FT MOD RES 310 310
FT MOD RES 311 311
FT MOD RES 312 312
FT MOD RES 313 313
FT MOD RES 314 314
FT MOD RES 315 315
FT MOD RES 316 316
FT MOD RES 317 317
FT MOD RES 318 318
FT MOD RES 319 319
FT MOD RES 320 320
FT MOD RES 321 321
FT MOD RES 322 322
FT MOD RES 323 323
FT MOD RES 324 324
FT MOD RES 325 325
FT MOD RES 326 326
FT MOD RES 327 327
FT MOD RES 328 328
FT MOD RES 329 329
FT MOD RES 330 330
FT MOD RES 331 331
FT MOD RES 332 332
FT MOD RES 333 333
FT MOD RES 334 334
FT MOD RES 335 335
FT MOD RES 336 336
FT MOD RES 337 337
FT MOD RES 338 338
FT MOD RES 339 339
FT MOD RES 340 340
FT MOD RES 341 341
FT MOD RES 342 342
FT MOD RES 343 343
FT MOD RES 344 344
FT MOD RES 345 345
FT MOD RES 346 346
FT MOD RES 347 347
FT MOD RES 348 348
FT MOD RES 349 349
FT MOD RES 350 350
FT MOD RES 351 351
FT MOD RES 352 352
FT MOD RES 353 353
FT MOD RES 354 354
FT MOD RES 355 355
FT MOD RES 356 356
FT MOD RES 357 357
FT MOD RES 358 358
FT MOD RES 359 359
FT MOD RES 360 360
FT MOD RES 361 361
FT MOD RES 362 362
FT MOD RES 363 363
FT MOD RES 364 364
FT MOD RES 365 365
FT MOD RES 366 366
FT MOD RES 367 367
FT MOD RES 368 368
FT MOD RES 369 369
FT MOD RES 370 370
FT MOD RES 371 371
FT MOD RES 372 372
FT MOD RES 373 373
FT MOD RES 374 374
FT MOD RES 375 375
FT MOD RES 376 376
FT MOD RES 377 377
FT MOD RES 378 378
FT MOD RES 379 379
FT MOD RES 380 380
FT MOD RES 381 381
FT MOD RES 382 382
FT MOD RES 383 383
FT MOD RES 384 384
FT MOD RES 385 385
FT MOD RES 386 386
FT MOD RES 387 387
FT MOD RES 388 388
FT MOD RES 389 389
FT MOD RES 390 390
FT MOD RES 391 391
FT MOD RES 392 392
FT MOD RES 393 393
FT MOD RES 394 394
FT MOD RES 395 395
FT MOD RES 396 396
FT MOD RES 397 397
FT MOD RES 398 398
FT MOD RES 399 399
FT MOD RES 400 400
FT MOD RES 401 401
FT MOD RES 402 402
FT MOD RES 403 403
FT MOD RES 404 404
FT MOD RES 405 405
FT MOD RES 406 406
FT MOD RES 407 407
FT MOD RES 408 408
FT MOD RES 409 409
FT MOD RES 410 410
FT MOD RES 411 411
FT MOD RES 412 412
FT MOD RES 413 413
FT MOD RES 414 414
FT MOD RES 415 415
FT MOD RES 416 416
FT MOD RES 417 417
FT MOD RES 418 418
FT MOD RES 419 419
FT MOD RES 420 420
FT MOD RES 421 421
FT MOD RES 422 422
FT MOD RES 423 423
FT MOD RES 424 424
FT MOD RES 425 425
FT MOD RES 426 426
FT MOD RES 427 427
FT MOD RES 428 428
FT MOD RES 429 429
FT MOD RES 430 430
FT MOD RES 431 431
FT MOD RES 432 432
FT MOD RES 433 433
FT MOD RES 434 434
FT MOD RES 435 435
FT MOD RES 436 436
FT MOD RES 437 437
FT MOD RES 438 438
FT MOD RES 439 439
FT MOD RES 440 440
FT MOD RES 441 441
FT MOD RES 442 442
FT MOD RES 443 443
FT MOD RES 444 444
FT MOD RES 445 445
FT MOD RES 446 446
FT MOD RES 447 447
FT MOD RES 448 448
FT MOD RES 449 449
FT MOD RES 450 450
FT MOD RES 451 451
FT MOD RES 452 452
FT MOD RES 453 453
FT MOD RES 454 454
FT MOD RES 455 455
FT MOD RES 456 456
FT MOD RES 457 457
FT MOD RES 458 458
FT MOD RES 459 459
FT MOD RES 460 460
FT MOD RES 461 461
FT MOD RES 462 462
FT MOD RES 463 463
FT MOD RES 464 464
FT MOD RES 465 465
FT MOD RES 466 466
FT MOD RES 467 467
FT MOD RES 468 468
FT MOD RES 469 469
FT MOD RES 470 470
FT MOD RES 471 471
FT MOD RES 472 472
FT MOD RES 473 473
FT MOD RES 474 474
FT MOD RES 475 475
FT MOD RES 476 476
FT MOD RES 477 477
FT MOD RES 478 478
FT MOD RES 479 479
FT MOD RES 480 480
FT MOD RES 481 481
FT MOD RES 482 482
FT MOD RES 483 483
FT MOD RES 484 484
FT MOD RES 485 485
FT MOD RES 486 486
FT MOD RES 487 487
FT MOD RES 488 488
FT MOD RES 489 489
FT MOD RES 490 490
FT MOD RES 491 491
FT MOD RES 492 492
FT MOD RES 493 493
FT MOD RES 494 494
FT MOD RES 495 495
FT MOD RES 496 496
FT MOD RES 497 497
FT MOD RES 498 498
FT MOD RES 499 499
FT MOD RES 500 500
FT MOD RES 501 501
FT MOD RES 502 502
FT MOD RES 503 503
FT MOD RES 504 504
FT MOD RES 505 505
FT MOD RES 506 506
FT MOD RES 507 507
FT MOD RES 508 508
FT MOD RES 509 509
FT MOD RES 510 510
FT MOD RES 511 511
FT MOD RES 512 512
FT MOD RES 513 513
FT MOD RES 514 514
FT MOD RES 515 515
FT MOD RES 516 516
FT MOD RES 517 517
FT MOD RES 518 518
FT MOD RES 519 519
FT MOD RES 520 520
FT MOD RES 521 521
FT MOD RES 522 522
FT MOD RES 523 523
FT MOD RES 524 524
FT MOD RES 525 525
FT MOD RES 526 526
FT MOD RES 527 527
FT MOD RES 528 528
FT MOD RES 529 529
FT MOD RES 530 530
FT MOD RES 531 531
FT MOD RES 532 532
FT MOD RES 533 533
FT MOD RES 534 534
FT MOD RES 535 535
FT MOD RES 536 536
FT MOD RES 537 537
FT MOD RES 538 538
FT MOD RES 539 539
FT MOD RES 540 540
FT MOD RES 541 541
FT MOD RES 542 542
FT MOD RES 543 543
FT MOD RES 544 544
FT MOD RES 545 545
FT MOD RES 546 546
FT MOD RES 547 547
FT MOD RES 548 548
FT MOD RES 549 549
FT MOD RES 550 550
FT MOD RES 551 551
FT MOD RES 552 552
FT MOD RES 553 553
FT MOD RES 554 554
FT MOD RES 555 555
FT MOD RES 556 556
FT MOD RES 557 557
FT MOD RES 558 558
FT MOD RES 559 559
FT MOD RES 560 560
FT MOD RES 561 561
FT MOD RES 562 562
FT MOD RES 563 563
FT MOD RES 564 564
FT MOD RES 565 565
FT MOD RES 566 566
FT MOD RES 567 567
FT MOD RES 568 568
FT MOD RES 569 569
FT MOD RES 570 570
FT MOD RES 571 571
FT MOD RES 572 572
FT MOD RES 573 573
FT MOD RES 574 574
FT MOD RES 575 575
FT MOD RES 576 576
FT MOD RES 577 577
FT MOD RES 578 578
FT MOD RES 579 579
FT MOD RES 580 580
FT MOD RES 581 581
FT MOD RES 582 582
FT MOD RES 583 583
FT MOD RES 584 584
FT MOD RES 585 585
FT MOD RES 586 586
FT MOD RES 587 587
FT MOD RES 588 588
FT MOD RES 589 589
FT MOD RES 590 590
FT MOD RES 591 591
FT MOD RES 592 592
FT MOD RES 593 593
FT MOD RES 594 594
FT MOD RES 595 595
FT MOD RES 596 596
FT MOD RES 597 597
FT MOD RES 598 598
FT MOD RES 599 599
FT MOD RES 600 600
FT MOD RES 601 601
FT MOD RES 602 602
FT MOD RES 603 603
FT MOD RES 604 604
FT MOD RES 605 605
FT MOD RES 606 606
FT MOD RES 607 607
FT MOD RES 608 608
FT MOD RES 609 609
FT MOD RES 610 610
FT MOD RES 611 611
FT MOD RES 612 612
FT MOD RES 613 613
FT MOD RES 614 614
FT MOD RES 615 615
FT MOD RES 616 616
FT MOD RES 617 617
FT MOD RES 618 618
FT MOD RES 619 619
FT MOD RES 620 620
FT MOD RES 621 621
FT MOD RES 622 622
FT MOD RES 623 623
FT MOD RES 624 624
FT MOD RES 625 625
FT MOD RES 626 626
FT MOD RES 627 627
FT MOD RES 628 628
FT MOD RES 629 629
FT MOD RES 630 630
FT MOD RES 631 631
FT MOD RES 632 632
FT MOD RES 633 633
FT MOD RES 634 634
FT MOD RES 635 635
FT MOD RES 636 636
FT MOD RES 637 637
FT MOD RES 638 638
FT MOD RES 639 639
FT MOD RES 640 640
FT MOD RES 641 641
FT MOD RES 642 642
FT MOD RES 643 643
FT MOD RES 644 644
FT MOD RES 645 645
FT MOD RES 646 646
FT MOD RES 647 647
FT MOD RES 648 648
FT MOD RES 649 649
FT MOD RES 650 650
FT MOD RES 651 651
FT MOD RES 652 652
FT MOD RES 653 653
FT MOD RES 654 654
FT MOD RES 655 655
FT MOD RES 656 656
FT MOD RES 657 657
FT MOD RES 658 658
FT MOD RES 659 659
FT MOD RES 660 660
FT MOD RES 661 661
FT MOD RES 662 662
FT MOD RES 663 663
FT MOD RES 664 664
FT MOD RES 665 665
FT MOD RES 666 666
FT MOD RES 667 667
FT MOD RES 668 668
FT MOD RES 669 669
FT MOD RES 670 670
FT MOD RES 671 671
FT MOD RES 672 672
FT MOD RES 673 673
FT MOD RES 674 674
FT MOD RES 675 675
FT MOD RES 676 676
FT MOD RES 677 677
FT MOD RES 678 678
FT MOD RES 679 679
FT MOD RES 680 680
FT MOD RES 681 681
FT MOD RES 682 682
FT MOD RES 683 683
FT MOD RES 684 684
FT MOD RES 685 685
FT MOD RES 686 686
FT MOD RES 687 687
FT MOD RES 688 688
FT MOD RES 689 689
FT MOD RES 690 690
FT MOD RES 691 691
FT MOD RES 692 692
FT MOD RES 693 693
FT MOD RES 694 694
FT MOD RES 695 695
FT MOD RES 696 696
FT MOD RES 697 697
FT MOD RES 698 698
FT MOD RES 699 699
FT MOD RES 700 700
FT MOD RES 701 701
FT MOD RES 702 702
FT MOD RES 703 703
FT MOD RES 704 704
FT MOD RES 705 705
FT MOD RES 706 706
FT MOD RES 707 707
FT MOD RES 708 708
FT MOD RES 709 709
FT MOD RES 710 710
FT MOD RES 711 711
FT MOD RES 712 712
FT MOD RES 713 713
FT MOD RES 714 714
FT MOD RES 715 715
FT MOD RES 716 716
FT MOD RES 717 717
FT MOD RES 718 718
FT MOD RES 719 719
FT MOD RES 720 720
FT MOD RES 721 721
FT MOD RES 722 722
FT MOD RES 723 723
FT MOD RES 724 724
FT MOD RES 725 725
FT MOD RES 726 726
FT MOD RES 727 727
FT MOD RES 728 728
FT MOD RES 729 729
FT MOD RES 730 730
FT MOD RES 731 731
FT MOD RES 732 732
FT MOD RES 733 733
FT MOD RES 734 734
FT MOD RES 735 735
FT MOD RES 736 736
FT MOD RES 737 737
FT MOD RES 738 738
FT MOD RES 739 739
FT MOD RES 740 740
FT MOD RES 741 741
FT MOD RES 742 742
FT MOD RES 743 743
FT MOD RES 744 744
FT MOD RES 745 745
FT MOD RES 746 746
FT MOD RES 747 747
FT MOD RES 748 748
FT MOD RES 749 749
FT MOD RES 750 750
FT MOD RES 751 751
FT MOD RES 752 752
FT MOD RES 753 753
FT MOD RES 754 754
FT MOD RES 755 755
FT MOD RES 756 756
FT MOD RES 757 757
FT MOD RES 758 758
FT MOD RES 759 759
FT MOD RES 760 760
FT MOD RES 761 761
FT MOD RES 762 762
FT MOD RES 763 763
FT MOD RES 764 764
FT MOD RES 765 765
FT MOD RES 766 766
FT MOD RES 767 767
FT MOD RES 768 768
FT MOD RES 769 769
FT MOD RES 770 770
FT MOD RES 771 771
FT MOD RES 772 772
FT MOD RES 773 773
FT MOD RES 774 774
FT MOD RES 775 775
FT MOD RES 776 776
FT MOD RES 777 777
FT MOD RES 778 778
FT MOD RES 779 779
FT MOD RES 780 780
FT MOD RES 781 781
FT MOD RES 782 782
FT MOD RES 783 783
FT MOD RES 784 784
FT MOD RES 785 785
FT MOD RES 786 786
FT MOD RES 787 787
FT MOD RES 788 788
FT MOD RES 789 789
FT MOD RES 790 790
FT MOD RES 791 791
FT MOD RES 792 792
FT MOD RES 793 793
FT MOD RES 794 794
FT MOD RES 795 795
FT MOD RES 796 796
FT MOD RES 797 797
FT MOD RES 798 798
FT MOD RES 799 799
FT MOD RES 800 800
FT MOD RES 801 801
FT MOD RES 802 802
FT MOD RES 803 803
FT MOD RES 804 804
FT MOD RES 805 805
FT MOD RES 806 806
FT MOD RES 807 807
FT MOD RES 808 808
FT MOD RES 809 809
FT MOD RES 810 810
FT MOD RES 811 811
FT MOD RES 812 812
FT MOD RES 813 813
FT MOD RES 814 814
FT MOD RES 815 815
FT MOD RES 816 816
FT MOD RES 817 817
FT MOD RES 818 818
FT MOD RES 819 819
FT MOD RES 820 820
FT MOD RES 821 821
FT MOD RES 822 822
FT MOD RES 823 823
FT MOD RES 824 824
FT MOD RES 825 825
FT MOD RES 826 826
FT MOD RES 827 827
FT MOD RES 828 828
FT MOD RES 829 829
FT MOD RES 830 830
FT MOD RES 831 831
FT MOD RES 832 832
FT MOD RES 833 833
FT MOD RES 834 834
FT MOD RES 835 835
FT MOD RES 836 836
FT MOD RES 837 837
FT MOD RES 838 838
FT MOD RES 839 839
FT MOD RES 840 840
FT MOD RES 841 841
FT MOD RES 842 842
FT MOD RES 843 843
FT MOD RES 844 844
FT MOD RES 845 845
FT MOD RES 846 846
FT MOD RES 847 847
FT MOD RES 848 848
FT MOD RES 849 849
FT MOD RES 850 850
FT MOD RES 851 851
FT MOD RES 852 852
FT MOD RES 853 853
FT MOD RES 854 854
FT MOD RES 855 855
FT MOD RES 856 856
FT MOD RES 857 857
FT MOD RES 858 858
FT MOD RES 859 859
FT MOD RES 860 860
FT MOD RES 861 861
FT MOD RES 862 862
FT MOD RES 863 863
FT MOD RES 864 864
FT MOD RES 865 865
FT MOD RES 866 866
FT MOD RES 867 867
FT MOD RES 868 868
FT MOD RES 869 869
FT MOD RES 870 870
FT MOD RES 871 871
FT MOD RES 872 872
FT MOD RES 873 873
FT MOD RES 874 874
FT MOD RES 875 875
FT MOD RES 876 876
FT MOD RES 877 877
FT MOD RES 878 878
FT MOD RES 879 879
FT MOD RES 880 880
FT MOD RES 881 881
FT MOD RES 882 882
FT MOD RES 883 883
FT MOD RES 884 884
FT MOD RES 885 885
FT MOD RES 886 886
FT MOD RES 887 887
FT MOD RES 888 888
FT MOD RES 889 889
FT MOD RES 890 890
FT MOD RES 891 891
FT MOD RES 892 892
FT MOD RES 893 893
FT MOD RES 894 894
FT MOD RES 895 895
FT MOD RES 896 896
FT MOD RES 897 897
FT MOD RES 898 898
FT MOD RES 899 899
FT MOD RES 900 900
FT MOD RES 901 901
FT MOD RES 902 902
FT MOD RES 903 903
FT MOD RES 904 904
FT MOD RES 905 905
FT MOD RES 906 906
FT MOD RES 907 907
FT MOD RES 908 908
FT MOD RES 909 909
FT MOD RES 910 910
FT MOD RES 911 911
FT MOD RES 912 912
FT MOD RES 913 913
FT MOD RES 914 914
FT MOD RES 915 915
FT MOD RES 916 916
FT MOD RES 917 917
FT MOD RES 918 918
FT MOD RES 919 919
FT MOD RES 920 920
FT MOD RES 921 921
FT MOD RES 922 922
FT MOD RES 923 923
FT MOD RES 924 924
FT MOD RES 925 925
FT MOD RES 926 926
FT MOD RES 927 927
FT MOD RES 928 928
FT MOD RES 929 929
FT MOD RES 930 930
FT MOD RES 931 931
FT MOD RES 932 932
FT MOD RES 933 933
FT MOD RES 934 934
FT MOD RES 935 935
FT MOD RES 936 936
FT MOD RES 937 937
FT MOD RES 938 938
FT MOD RES 939 939
FT MOD RES 940 940
FT MOD RES 941 941
FT MOD RES 942 942
FT MOD RES 943 943
FT MOD RES 944 944
FT MOD RES 945 945
FT MOD RES 946 946
FT MOD RES 947 947
FT MOD RES 948 948
FT MOD RES 949 949
FT MOD RES 950 950
FT MOD RES 951 951
FT MOD RES 952 952
FT MOD RES 953 953
FT MOD RES 954 954
FT MOD RES 955 955
FT MOD RES 956 956
FT MOD RES 957 957
FT MOD RES 958 958
FT MOD RES 959 959
FT MOD RES 960 960
FT MOD RES 961 961
FT MOD RES 962 962
FT MOD RES 963 963
FT MOD RES 964 964
FT MOD RES 965 965
FT MOD RES 966 966
FT MOD RES 967 967
FT MOD RES 968 968
FT MOD RES 969 969
FT MOD RES 970 970
FT MOD RES 971 971
FT MOD RES 972 972
FT MOD RES 973 973
FT MOD RES 974 974
FT MOD RES 975 975
FT MOD RES 976 976
FT MOD RES 977 977
FT MOD RES 978 978
FT MOD RES 979 979
FT MOD RES 980 980
FT MOD RES 981 981
FT MOD RES 982 982
FT MOD RES 983 983
FT MOD RES 984 984
FT
```

RESULT 2

```

RN3_YEAST      STANDARD;      PRT;      627 AA.
ID   RN3_YEAST
AC   P36070.1
DT   01-JUN-1994 (Rel. 29, Created)
DR   01-JUN-1994 (Rel. 29, Last sequence update)
DE   16-OCT-2001 (Rel. 40, Last annotation update)
DN   RNA polymerase I specific transcription initiation factor RN3.
GN   RN3 OR YKL125W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=96324404; PubMed=8670901;
RA   Yamamoto R.T., Nogai Y., Dodd J.A., Nomura M.;
RT   "RN3 gene of Saccharomyces cerevisiae encodes an essential RNA
RT   polymerase I transcription factor which interacts with the polymerase
RT   independently of DNA template.";
RL   EMBL J. 15:364-3973(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rad M.R., Xu G., Kirchraeth L., Fritz C., Keuchel H., Hollenberg C.P.;
RL   Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC   -1- FUNCTION: REQUIRED FOR EFFICIENT TRANSCRIPTION INITIATION BY RNA
CC   POLYMERASE I. INTERACTS WITH POL I IN THE ABSENCE OF TEMPLATE DNA
CC   AND STIMULATES RECRUITMENT OF POL I, BUT DOES NOT REMAIN AS PART
CC   OF STABLE PRE-INITIATION COMPLEX.
CC   -1- SUBUNIT: Monomer (Probable).
CC   -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -1- SIMILARITY: TO S.POMBE SPAC1866.11C AND C.ELEGANS C36E8.1.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; Z71927; CAA96470.1; -
DR   EMBL; Z28125; CAA81966.1; -
DR   PIR; S37954; S37954.
DR   TRANSPAC; T03565; -
DR   SGD; S0001608; RN3.
DR   Pfam; PF05327; RN3.1.
KW   Transcription regulation; Nuclear protein.
FT   DOMAIN 252..259
FT   DOMAIN 267..274
FT   DOMAIN 277..280
FT   DOMAIN 546..549
SQ   SEQUENCE 627 AA; 72387 MW; A31E7386A987FDB CRC64;

```

Query Match 89.5%; Score 34; DB 1; Length 627;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SINFEX 7
 DB 118 SINFEX 124

RESULT 3
 OVAL_COTUA STANDARD; PRT; 382 AA.
 ID OVAL_COTUA
 AC P19104.1
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Ovalbumin.

```

OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX   NCBI_TaxID=93934;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oviduct;
RX   MEDLINE=91016850; PubMed=2216734;
RA   Mucha J., Klaudivova V., Hanes J., Simuth J.;
RT   "The sequence of Japanese quail ovalbumin cDNA.";
RL   Nucleic Acids Res. 18:5553-5553 (1990).
CC   -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X53964; CAA37916.1; -
DR   PIR; S11433; S11433.
DR   HSSP; P01012; IOVA.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00079; serpin; 1.
DR   SMART; SM00093; SERPIN; 1.
DR   PROSITE; PS00284; SERPIN; 1.
KW   Serpin; Glycoprotein.
FT   INIT MET 0
FT   DISULFID 73..120
FT   CARBOHYD 292..292
FT   CARBOHYD 311..311
FT   ACT SITE 352..353
SQ   SEQUENCE 382 AA; 42108 MW; A4B859A1BA8EF316 CRC64;

```

Query Match 86.8%; Score 33; DB 1; Length 382;
 Best Local Similarity 87.5%; Pred. No. 11;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SINFEX 8
 DB 257 SINFEX 264

RESULT 4
 RM09_HUMAN STANDARD; PRT; 267 AA.
 ID RM09_HUMAN
 AC O98YD2; O9BSW8; 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 60S ribosomal protein L9, mitochondrial precursor (L9mt).
 GN MRP19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21293042; PubMed=11279069;
 RA Suzuki T., Terasaki M., Takemoto-Hori C., Hanada T., Ueda T., Wada A.,
 RT "Structural compensation for the deficit of rRNA with proteins in the
 RT mammalian mitochondrial ribosome. Systematic analysis of protein
 RT components of the large ribosomal subunit from mammalian
 RT mitochondria.";
 RL J. Biol. Chem. 276:21724-21736 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;

```

RA Strausberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buerger K.H., Scheffer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Locantello N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M.,
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE L9P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB049636; BAB40841.1; -
DR EMBL; BC004517; AAH04517.1; -
DR Genew; HGNC:14277; MRPL9.
DR InterPro; IPR000244; Ribosomal_L9.
DR Pfam; PF01281; Ribosomal_L9_N; 1.
DR Ribosomal protein; Mitochondrion; Transic peptide.
FT CHAIN 1 267 MITOCHONDRION (POTENTIAL).
FT TRANSIT 1 267 60S RIBOSOMAL PROTEIN L9.
FT CONFLICT 210 210 A -> E (TN REF. 2).
FT SEQUENCE 267 AA; 30185 MW; 346C254220FDB4 CRC64;
SQ
Query Match 84.2%; Score 32; DB 1; Length 267;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 STINPEK 7
DB 236 SVINPEK 242

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20509; AAA62167.1; -
DR HSP; P00943; 2BTM.
DR HAMAP; MF_00147; -; 1.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR TIGRPFAMs; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
DR Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KM Pentose shunt.
FT ACT SITE 98 98 BY SIMILARITY.
FT ACT SITE 167 167 BY SIMILARITY.
FT SEQUENCE 242 AA; 26969 MW; B1E8560E2DA18FA1 CRC64;
SQ
Query Match 81.6%; Score 31; DB 1; Length 242;
Best Local Similarity 71.4%; Pred. No. 18;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 STINPEK 7
DB 155 SVINPEK 161

```

```

RESULT 6
YXCD CYAPA
ID YXCD CYAPA STANDARD; PRT; 244 AA.
AC P48334;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable ABC transporter ATP-binding protein in yef23-apef intergenic
DE region (ORF244).
OS Cyanophora paradoxa.
OG Cyanella.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanella DNA from Cyanophora paradoxa.";
RT Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Neumann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete complexity of the cyanella genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RT (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmer W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; U30821; AAA81304.1; -
DR PIR; T06961; T06961.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR Hypothetical protein; ATP-binding; Transport; Cyanelle.
FT NP_BIND 41 48 ATP (POTENTIAL).
SQ SEQUENCE 244 AA; 27747 MW; 4CSB357F9C55D3B CRC64;

Query Match
Best Local Similarity 81.6%; Score 31; DB 1; Length 244;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 STINFEKT 8
DB 204 STINFEKT 211

RESULT 7
MTH5_HAEIN STANDARD; PRT; 304 AA.
AC P45000;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Modification methylase Hindv (EC 2.1.1.73) (Cytosine-specific
  methyltransferase Hindv) (M.Hindv).
GN HINDVM OR H11041.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
  Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
  Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
  Weiman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
  Utechtback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
  Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
  Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
  Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
  Rd.";
RT Science 269:496-512(1995).
RL
CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
  GRCGCG, CAUSES SPECIFIC METHYLATION ON C-? ON BOTH STRANDS, AND
  PROTECTS THE DNA FROM CLEAVAGE BY THE HINDV ENDONULEASE.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
  adenosyl-L-homocysteine + DNA 5-methylcytosine.
CC -1- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32784; AAC22700.1; -
DR PIR; C64109; C64109.
DR HSSP; O14717; 1G55.
DR REBASE; 3574; M.Hindv.
DR TIGR; H11041; -

```

```

DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF00145; DNA_methylase; 1.
DR PRINTS; PR00105; CSMETTRFRASE.
DR TIGRPFAM; TIGR00675; dcm; 1.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00095; C5_MTASE_2; 1.
DR Hypothetical protein; Transferase; Methyltransferase;
  Restriction system; Complete proteome.
KW ACT SITE 75 75
FT ACT_SITE 75 75
SQ SEQUENCE 304 AA; 34365 MW; 03DA1EAB27C84BBD CRC64;

Query Match
Best Local Similarity 81.6%; Score 31; DB 1; Length 304;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 STINFEKT 7
DB 291 ATINFEKT 297

RESULT 8
R11_BUCAP STANDARD; PRT; 231 AA.
ID R11_BUCAP
AC Q8KA67;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L1.
GN RPLA OR BUSG038.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klaesson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
  Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
  LOCATED IN THE NEIGHBORHOOD OF THE SITE WHERE ELONGATION FACTOR TU
  IS BOUND TO THE RIBOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014080; AAM67609.1; -
DR InterPro; IPR005878; L1_bact_chl.
DR InterPro; IPR002143; Ribosomal_L1.
DR Pfam; PF00687; Ribosomal_L1; 1.
DR ProDom; PD001314; Ribosomal_L1; 1.
DR TIGRPFAM; TIGR01169; rplA_bact; 1.
DR PROSITE; PS01199; RIBOSOMAL_L1; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 231 AA; 25609 MW; 4D334DACF32A3C2 CRC64;

Query Match
Best Local Similarity 78.9%; Score 30; DB 1; Length 231;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 INFEKT 8
DB 15 INFEKT 20

RESULT 9

```

OVAX CHICK STANDARD; PRT; 232 AA.
 ID OVAX CHICK
 AC P01013;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Gene X protein (Ovalbumin-related) (Fragment).
 GN X.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81022623; PubMed=7418002;
 RA Helling R., Petrin F., Gannon F., Mandel J.L., Chambon P.;
 RT "The ovalbumin gene family: structure of the X gene and evolution of
 RT Cell 20:625-637(1980).
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OY-SERPIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; J00920; AAA68881.1; -
 DR EMBL; J00918; AAA68881.1; JOINED.
 DR EMBL; J00919; AAA68881.1; JOINED.
 DR EMBL; V00385; CAA23683.1; -
 DR EMBL; V00386; CAA23684.1; -
 DR EMBL; V00387; CAA23685.1; -
 DR PIR; A01243; DXCH.
 DR HSSP; P01012; IOVA.
 DR InterPro; IPR000215; Serpin.
 DR Pfam; PF00079; serpin_1.
 DR SMART; SM00093; SERPIN; 1.
 DR PROSITE; PS00284; SERPIN; 1.
 KW Serpin.
 FT NON TER
 SQ SEQUENCE 232 AA; 26291 MW; 6B5B86EC4D3B9195 CRC64;
 Query Match 78.9%; Score 30; DB 1; Length 232;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 INFECL 8
 DB 104 INFECL 109
 RESULT 10
 ID Y567_HELPJ STANDARD; PRT; 349 AA.
 AC Q9ZLR4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein JHP0514.
 GN JHP0514.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923662;
 RA Alm R.A., Ling L.-S.L., Moir D.T., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori."
 RL Nature 397:176-180(1999).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AE001484; AAD06090.1; -
 DR PIR; H71923; H71923.
 DR InterPro; IPR002549; UPF0118.
 DR Pfam; PF01594; UPF0118; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 56 76 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 195 215 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 300 320 POTENTIAL.
 SQ SEQUENCE 349 AA; 39804 MW; A1846D48CB3A86F CRC64;
 Query Match 78.9%; Score 30; DB 1; Length 349;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 INFECL 8
 DB 85 INFECL 90
 RESULT 11
 ID SERC DROME STANDARD; PRT; 364 AA.
 AC Q9VAN0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).
 GN ESTS:39C108 OR CG11899.
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champ M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKlos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu L.B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Garbalian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Katush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuhl D., Lai Z.,
RA Laekso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Many D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reilbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- CATALYTIC ACTIVITY: O-phospho-L-serine + 2-oxoglutarate = 3-
CC phosphonooxypruvate + L-glutamate.
CC -1- COFACTOR: Pyridoxal phosphate.
CC -1- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE
CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.
CC -1- SIMILARITY: Belongs to class-V of pyridoxal-phosphate-dependent
CC aminotransferases.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AE003768; AAF56874.1; -.
CC DR HSSP; P23721; IBJN.
CC DR FLYBASE; FBgn0014427; ESTS:39C10S.
CC DR InterPro; IPR000192; Aminotransf.
CC DR InterPro; IPR003248; Pser aminotransf.
CC DR Pfam; PF00266; aminotran_5; 1.
CC DR ProDom; PD001544; Pser_aminotransf; 1.
CC DR TIGRPFAMS; TIGR01364; serc_1; 1.
CC DR PROSITE; PS00595; AA_TRANSFERS CLASS 5; 1.
CC KW Serine biosynthesis; Transferase; Aminotransferase;
CC PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC FT BINDING 194
CC SEQUENCE 364 AA; 39540 MW; DAEN4E2F5BD4DB74 CRC64;
SQ
Query Match 78.9%; Score 30; DB 1; Length 364;
Best Local Similarity 62.5%; Pred. No. 46;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEXL 8
DB 220 SIINFEXOM 227

```

```

OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCB1_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapteval V., Anderson I., Ivanova N., Reznik G., Los T., Lykdis A.,
RA Bhattacharya A., Bartman A., Gardner W., Gueckin G., Zhu L.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fontein M., Kyriades N., Overbeek R.;
RA "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol 184:2005-2018(2002).
CC -1- CATALYTIC ACTIVITY: ATP + oxaloacetate = ADP + phosphoenolpyruvate
CC + CO(2).
CC -1- PATHWAY: Rate-limiting gluconeogenic enzyme.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: Belongs to the phosphoenolpyruvate carboxykinase (ATP)
CC family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AE010616; AAL95316.1; -.
CC DR HAMAP; MF_00453; -.
CC DR InterPro; IPR001272; PEPCCK_ATP.
CC DR Pfam; PF01293; PEPCCK_ATP; 1.
CC DR ProDom; PD004723; PEPCCK_ATP; 1.
CC DR TIGRPFAMS; TIGR00224; pcca; 1.
CC DR PROSITE; PS00532; PEPCCK_ATP; 1.
CC KW Gluconeogenesis; Lyase; Decarboxylase; ATP-binding; Complete proteome.
CC FT NP_BIND 230
CC SEQUENCE 527 AA; 59055 MW; 275849FDE254AC01 CRC64;
SQ
Query Match 78.9%; Score 30; DB 1; Length 527;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 IINFEX 7
DB 173 IINFEX 178

```

```

RESULT 12
PCK_FUSNN STANDARD; PRT; 527 AA.
AC Q8RE12;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49) (PEP
DE carboxykinase) (Phosphoenolpyruvate carboxylase) (PEPCK).
GN PCKA OR FN1120.

```

```

RESULT 13
MNT2 YEAST STANDARD; PRT; 558 AA.
AC P53059;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-1,3-mannosyltransferase MNT2 (EC 2.4.1.-).
GN MNT2 OR YGL257C OR NR0558.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCB1_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97127827; PubMed=8972578;
RA Colasac E., Maillier E., Robineau S., Netter P.;
RT "Sequence of a 39,411 bp DNA fragment covering the left end of
RL Yeast 12:1555-1562(1996).

```



```

RT bacterium Clostridium acetobutylicum.;
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE007771; AKK80682.1; -.
CC DR PIR; G97236; G97236.
CC DR InterPro; IPR003439; ABC transporter.
CC KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
CC DNA recombination; ATP-binding; Coiled coil; Complete proteome.
CC FT NP_BIND 35 42 ATP (POTENTIAL).
CC FT DOMAIN 197 415 COILED COIL (POTENTIAL).
CC FT DOMAIN 446 1003 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 1163 AA; 135507 MW; CE5F0BD2215D7A92 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 1163;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 INFEKL 8
Db 20 INFEKL 25

```

Search completed: January 30, 2004, 07:07:49
Job time : 14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:27 ; Search time 70 Seconds
(without alignments)
29.492 Million cell updates/sec

Title: SEQ10
Perfect score: 38
Sequence: 1 sinfekl 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	86.8	168	10 Q98RQ3	Q98RQ3 guillardia
2	33	86.8	315	10 Q65158	Q65158 phaseolus v
3	33	86.8	366	2 Q8VQ88	Q8VQ88 unclutred
4	32	84.2	260	16 Q9AA01	Q9AA01 caulobacter
5	32	84.2	299	16 Q92XR1	Q92XR1 rhizobium m
6	32	84.2	347	8 Q98RV6	Q98RV6 guillardia
7	32	84.2	443	17 Q8TRP5	Q8TRP5 mechanosarc
8	32	84.2	444	4 Q8IXU5	Q8IXU5 homo sapien
9	32	84.2	580	4 Q92565	Q92565 homo sapien
10	32	84.2	612	11 Q8BJT9	Q8BJT9 mus musculu
11	32	84.2	814	11 Q8CORS	Q8CORS mus musculu
12	32	84.2	814	11 Q8CQ09	Q8CQ09 mus musculu
13	32	84.2	965	2 Q9S4D1	Q9S4D1 staphylococ
14	32	84.2	965	2 Q8VVR0	Q8VVR0 staphylococ
15	32	84.2	1311	5 Q95NL8	Q95NL8 caenorhabdi
16	32	84.2	1347	5 Q95WR8	Q95WR8 caenorhabdi

17	32	84.2	1470	5 Q21218	Q21218 caenorhabdi
18	32	84.2	1573	5 Q9VMF3	Q9VMF3 drosophila
19	32	84.2	1573	5 Q95V18	Q95V18 drosophila
20	31	81.6	97	12 Q8BDP4	Q8BDP4 reindeer pa
21	31	81.6	103	5 Q8IB53	Q8IB53 plasmodium
22	31	81.6	211	5 P91370	P91370 caenorhabdi
23	31	81.6	287	11 Q8R1R1	Q8R1R1 mus musculu
24	31	81.6	307	17 Q59582	Q59582 pyrococcus
25	31	81.6	309	16 Q98OR5	Q98OR5 mycoplasma
26	31	81.6	364	10 Q9SKB4	Q9SKB4 arabidopsis
27	31	81.6	461	11 Q8BZK9	Q8BZK9 mus musculu
28	31	81.6	680	12 Q9Y7Q6	Q9Y7Q6 ateline her
29	31	81.6	834	11 Q8R3E5	Q8R3E5 mus musculu
30	31	81.6	876	11 Q8VCC8	Q8VCC8 mus musculu
31	31	81.6	881	4 Q8WVNO	Q8WVNO homo sapien
32	31	81.6	881	4 Q95634	Q95634 homo sapien
33	31	81.6	881	4 Q95398	Q95398 homo sapien
34	31	81.6	884	11 Q921C8	Q921C8 fatuus norv
35	31	81.6	1089	5 Q81605	Q81605 plasmodium
36	31	81.6	1113	4 Q8TEA3	Q8TEA3 homo sapien
37	31	81.6	1138	11 Q8CHG7	Q8CHG7 mus musculu
38	31	81.6	1204	4 Q9UHV4	Q9UHV4 homo sapien
39	31	81.6	1391	4 Q8TEU6	Q8TEU6 homo sapien
40	31	81.6	1499	4 Q9Y4G8	Q9Y4G8 homo sapien
41	31	81.6	1509	4 Q96PC1	Q96PC1 homo sapien
42	31	81.6	1601	4 Q8TEU7	Q8TEU7 homo sapien
43	31	81.6	1601	4 Q8NT21	Q8NT21 homo sapien
44	31	81.6	1836	10 Q9LFX4	Q9LFX4 arabidopsis
45	31	81.6	1909	10 Q9LFX3	Q9LFX3 arabidopsis

ALIGNMENTS

RESULT 1					
ID	Q98RQ3	PRELIMINARY;	PRT;	168 AA.	
AC	Q98RQ3:				
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)				
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)				
DE	Hypothetical 20.2 kDa protein orf168 from chromosome 1.				
GN	ORF168.				
OS	Guillardia theta (Cryptomonas phi).				
OC	Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.				
OX	NCBI_TaxID=55529;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=2122349; PubMed=1132671;				
RA	Douglas S. Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T., Wu X., Reich M., Cavalier-Smith T., Maier U.G.;				
RT	"The highly reduced genome of an enslaved algal nucleus."				
RL	Nature 410:1091-1096 (2001).				
DR	EMBL; AF165818; AAK39893.1; -				
KW	Hypothetical protein.				
SQ	SEQUENCE 168 AA; 20185 MW; 2874CBD53028A3D CRC64;				
Query Match					
Best Local Similarity 86.8%; Score 33; DB 10; Length 168;					
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;					
QY	1 SINFEKL 8				
DB	108 NINFEKI 115				
RESULT 2					
ID	Q65158	PRELIMINARY;	PRT;	315 AA.	
AC	Q65158:				
DT	01-AUG-1998 (TREMBLrel. 07, Created)				
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				

```

DE 1-aminocyclopropane-1-carboxylic acid oxidase.
GN ACO1.
OS Phaseolus vulgaris (kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Taylor; TISSUE=leaf;
RA Pidgeon C.M., Reid D.M., Facchini P.J.;
RT "Light induced changes in ethylene production in Phaseolus vulgaris
RT cv. Taylor.";
RL Plant Physiol. 114:165-165(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Taylor; TISSUE=leaf;
RA Pidgeon C.M., Facchini P.J., Reid D.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053354; AAC12934.1; -.
DR InterPro; IPR005123; 2OG-FcEl1_Oxy.
DR Pfam; PF03171; 2OG-FcEl1_Oxy; 1.
SQ SEQUENCE 315 AA; 36141 MW; 6898811DF88E9B81 CRC64;

Query Match
Best Local Similarity 86.8%; Score 33; DB 10; Length 315;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 IINFEKL 8
Db 6 VINFEKL 12

RESULT 3
O8VO88 PRELIMINARY; PRT; 366 AA.
AC O8VO88;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)
DE Formyletrihydrofolate synthetase (Fragment).
OS uncultured environmental landfill bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=184235;
RN [1]
RP SEQUENCE FROM N.A.
RA Uz I., Lindner A.S., Rasche M.E., Townsend T.G., Ogram A.V.;
RT "Microbial Characterization of Landfill Cover Soils of Different
RT Ages.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF459462; AAL67505.1; -.
DR InterPro; IPR000559; Fmlethyd_synth.
DR Pfam; PF01268; FTHFS; 1.
FT NON TER 1 1
FT NON TER 366 366
SQ SEQUENCE 366 AA; 39816 MW; EB4685D501E97026 CRC64;

Query Match
Best Local Similarity 86.8%; Score 33; DB 2; Length 366;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 IINFEKL 8
Db 346 IVNFEKL 352

RESULT 4
O9AA01 PRELIMINARY; PRT; 260 AA.
AC O9AA01;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)

```

```

DE Hypothetical protein CC0810.
GN CC0810.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Lau M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro J., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005757; AAK22795.1; -.
DR TIGR; CC0810; -.
DR InterPro; IPR005269; Cons_hypoth730.
DR Pfam; PF03641; Lysine_decarbox; 1.
DR TIGRFAMs; TIGR00730; TIGR00730; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 260 AA; 28789 MW; E4CF85223B12EF1B CRC64;

Query Match
Best Local Similarity 84.2%; Score 32; DB 16; Length 260;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 IINFEKL 8
Db 215 SINFEAL 222

RESULT 5
O92KR1 PRELIMINARY; PRT; 299 AA.
AC O92KR1;
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Hypothetical protein R00434.
GN R00434 OR SMC01731.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreaou S., Gloux S.,
RA Godie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramberger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591783; CAC41871.1; -.
DR InterPro; IPR005269; Cons_hypoth730.
DR Pfam; PF03641; Lysine_decarbox; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 299 AA; 33440 MW; 9BD61375BE91F195 CRC64;

Query Match
Best Local Similarity 84.2%; Score 32; DB 16; Length 299;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 IINFEKL 8

```

Db 246 STINEAL 253

RESULT 6

ID 098RV6 PRELIMINARY; PRT; 347 AA.

AC 098RV6; 01-OCT-2001 (TRENBLrel. 18, Created)

DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Cyclin B.

CYCB.

OS Guillardia theta (Cryptomonas phi).

OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.

OX NCBI_Taxid=55529;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2123349; PubMed=1123671; Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,

RT "The highly reduced genome of an enslaved algal nucleus."

RL Nature 410:1091-1096(2001).

CC -1. SIMILARITY: BELONGS TO THE CYCLIN FAMILY.

DR EMBL; AF165818; AAK39844.1; -.

DR InterPro; IPR006670; Cyclin.

DR InterPro; IPR006671; Cyclin_N.

DR Pfam; PF00134; Cyclin; 1.

DR SMART; SM00385; CYCLIN; 1.

DR PROSITE; PS00292; CYCLINS; 1.

DR Cell cycle; Cell division; Cyclin.

SQ SEQUENCE 347 AA; 41616 MW; CC579909C403A04D CRC64;

Query Match

Best Local Similarity 84.2%; Score 32; DB 8; Length 347;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINEKL 8

Db 86 NVLNFEKL 93

RESULT 7

ID 08TRP5 PRELIMINARY; PRT; 443 AA.

AC 08TRP5; 01-UN-2002 (TRENBLrel. 21, Created)

DT 01-UN-2002 (TRENBLrel. 21, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Iron ABC transporter, solute-binding protein.

GN MA130.

OS Methanosarcina acetivorans.

OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;

OC Methanosarcinaceae; Methanosarcina.

OX NCBI_Taxid=2214;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=C2A / ATCC 35395 / DSM 2834;

RA MEDLINE=21929760; PubMed=11932238;

RA Galagan J.E., Nalysaum C., Roy A., Endrizzi M.G., MacDonald P.,

RA Fitzhugh W., Calvo S., Engels R., Smitrov S., Alnoor D., Brown A.,

RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,

RA Linton L., McKean P., McKernan K., Talamas J., Turrell A., Ye W.,

RA Zimmer A., Barber R.D., Cam I., Graham D.E., Grahame D.A., Guss A.M.,

RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,

RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,

RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,

RA Ferry J.G., Jarrell K.F., Jung H., Macario A.J.L., Paulsen I.,

RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,

RA Metcalf W.W., Birren B.,

RT "The genome of Methanosarcina acetivorans reveals extensive metabolic

and physiological diversity."

RL Genome Res. 12:532-542(2002).

DR EMBL; AE010779; AAK04551.1; -.

DR InterPro; IPR002491; Peripla_BP.

DR Pfam; PF01497; Peripla_BP_2; 1.

KW Complete proteome.

SQ SEQUENCE 443 AA; 49537 MW; 0D3F2198B20E86C9 CRC64;

Query Match

Best Local Similarity 84.2%; Score 32; DB 17; Length 443;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINEKL 8

Db 166 TIINYEKL 173

RESULT 8

ID 08IXU5 PRELIMINARY; PRT; 444 AA.

AC 08IXU5; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Similar to guanine nucleotide exchange factor for Rap1,

DE M-Ras-regulated GEF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Strausberg R.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC039203; AAH39203.1; -.

SQ SEQUENCE 444 AA; 52051 MW; ECEBDP820DDFF13B CRC64;

Query Match

Best Local Similarity 84.2%; Score 32; DB 4; Length 444;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINEKL 8

Db 381 NLVNEKL 388

RESULT 9

ID 092565 PRELIMINARY; PRT; 580 AA.

AC 092565; 01-FEB-1997 (TRENBLrel. 02, Created)

DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Hypothetical protein KIAA0277.

GN KIAA0277.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE=Brain;

RA MEDLINE=97191544; PubMed=9039502;

RA Taniguchi T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,

RA Nagase A., Kotani H., Miyajima N., Nomura N.;

RT "Prediction of the coding sequences of unidentified human genes. VI.

RT the coding sequences of 80 new genes (K1A0201-K1A0280) deduced by

RT analysis of cDNA clones from cell line KG-1 and brain."

RT DNA Res. 3:321-329(1996).

DR EMBL; D87467; BA013406.1; -.

DR InterPro; IPR000651; RasGEF.

DR InterPro; IPR001895; RasGEF_CDC25.

DR Pfam; PF00617; RasGEF; 1.

DR Pfam; PF00618; RasGEF; 1.

DR SMART; SM00147; RASGEF; 1.
 DR SMART; SM00229; RASGEF; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 580 AA; 67733 MW; 732FB7A11DFDA1C CRC64;

Query Match
 Best Local Similarity 84.2%; Score 32; DB 4; Length 580;
 Pred. No. 2.1e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINPEKL 8
 ::::|
 DB 517 NLVNFEXL 524

RESULT 10
 Q8BJ9 PRELIMINARY; PRT; 612 AA.
 ID Q8BJ9
 AC Q8BJ9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical guanine-nucleotide dissociation stimulators CDC25 family/guanine nucleotide exchange factor for Ras-like GTPases;
 DE N-terminal motif containing protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Body;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 DR EMBL; AK083591; BAC38963.1; -.
 KM Hypothetical protein.
 SQ SEQUENCE 612 AA; 70913 MW; DBD552DDFD392B6 CRC64;

Query Match
 Best Local Similarity 84.2%; Score 32; DB 11; Length 612;
 Pred. No. 2.2e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINPEKL 8
 ::::|
 DB 548 NLVNFEXL 555

RESULT 11
 Q8C0R5 PRELIMINARY; PRT; 814 AA.
 ID Q8C0R5
 AC Q8C0R5;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical guanine-nucleotide dissociation stimulators CDC25 family/guanine nucleotide exchange factor for Ras-like GTPases;
 DE N-terminal motif containing protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).

DR EMBL; AK029995; BAC26723.1; -.
 KM Hypothetical protein.
 SQ SEQUENCE 814 AA; 93681 MW; A64900A72AA3E89B CRC64;

Query Match
 Best Local Similarity 84.2%; Score 32; DB 11; Length 814;
 Pred. No. 2.9e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINPEKL 8
 ::::|
 DB 750 NLVNFEXL 757

RESULT 12
 Q8C0G9 PRELIMINARY; PRT; 814 AA.
 ID Q8C0G9
 AC Q8C0G9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical guanine-nucleotide dissociation stimulators CDC25 family/guanine nucleotide exchange factor for Ras-like GTPases;
 DE N-terminal motif containing protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 DR EMBL; AK030016; BAC26736.1; -.
 KM Hypothetical protein.
 SQ SEQUENCE 814 AA; 93754 MW; AA04BEC74CE13735 CRC64;

Query Match
 Best Local Similarity 84.2%; Score 32; DB 11; Length 814;
 Pred. No. 2.9e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINPEKL 8
 ::::|
 DB 750 NLVNFEXL 757

RESULT 13
 Q9S4D1 PRELIMINARY; PRT; 965 AA.
 ID Q9S4D1
 AC Q9S4D1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Lanthibiotic modifying enzyme.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 NCBI_TaxId=1280;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C55;
 RX MEDLINE=99346225; PubMed=10417203;
 RA Navaratna M.A., Sahl H.G., Tagg J.R.,
 RT "Identification of genes encoding two-component lanthibiotic production in Staphylococcus aureus C55 and other phage group II S. aureus strains and demonstration of an association with the exfoliative toxin B gene.";
 RT Infect. Immun. 67:4268-4271 (1999).
 DR EMBL; AF147744; AAD47013.1; -.
 SQ SEQUENCE 965 AA; 111570 MW; BC1EADABDAC4F346 CRC64;

Query Match 84.2%; Score 32; DB 2; Length 965;
 Best Local Similarity 71.4%; Pred. No. 3.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEX 7
 DB 202 SVNFEX 208

RESULT 14

OSVVR0 PRELIMINARY; PRT; 965 AA.

AC OSVVR0; 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

OS Staphylococcus aureus.

OG Plasmid ETB plasmid.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI_Taxid=1280;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TY4;

RX MEDLINE=21562640; PubMed=11705958;

RA Yamaguchi T., Hayashi T., Takami H., Ohnishi M., Murata T.,

RT Nakayama K., Asakawa K., Ohara M., Komatsuza H., Sugai M.,

RT "Complete Nucleotide Sequence of a Staphylococcus aureus Exfoliative

RT Toxin B Plasmid and Identification of a Novel ADP-Ribosyltransferase,

RT EDIN-C."

DR Infect. Immun. 69:7760-7771(2001).

DR EMBL; AP003088; BAB78440.1; -.

KW Plasmid.

SQ SEQUENCE 965 AA; 111561 MM; COBEPD8DACSF828 CRC64;

Query Match 84.2%; Score 32; DB 2; Length 965;

Best Local Similarity 71.4%; Pred. No. 3.4e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEX 7

DB 202 SVNFEX 208

RESULT 15

OS5N18 PRELIMINARY; PRT; 1311 AA.

AC OS5N18; 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE T14G10.2b protein (PKF isoform B).

GN T14G10.2 OR T14G10.2B.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_Taxid=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Wild A.;

RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology."

RL Science 283:2012-2018(1998).

RN [3]

RP SEQUENCE FROM N.A.

RA Verneijen M., van Berkel W., Jansen G., de Rooij J., Plasterk R.H.,

RA Bos J.L., Zwartkruis F.J.T.;

RT "Characterization of pxf, the C. elegans homolog of human PDZ-GFPs."

RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Wild A.;

RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

DR EMBL; 268880; CAC42342.1; -.

DR EMBL; 269664; CAC42342.1; JOINED.

DR EMBL; AF308448; AAL09434.1; -.

DR EMBL; 269664; CAC42313.1; -.

DR EMBL; 268880; CAC42313.1; JOINED.

DR WormPep; T14G10.2b; CE28081.

DR InterPro; IPR000595; CNMP_binding.

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR000651; RASGEFN.

DR InterPro; IPR001895; RASGEF_CDC25.

DR Pfam; PF00027; CNMP_binding; 2.

DR Pfam; PF00595; PDZ; 1.

DR Pfam; PF00789; RA; 1.

DR Pfam; PF00617; RASGEF; 1.

DR SMART; SM00100; CNMP; 1.

DR SMART; SM00228; PDZ; 1.

DR SMART; SM00314; RA; 1.

DR SMART; SM00147; RASGEF; 1.

DR SMART; SM00229; RASGEFN; 1.

DR PROSITE; PS50042; CNMP_BINDING_3; 1.

DR PROSITE; PS50106; PDZ; 1.

SQ SEQUENCE 1311 AA; 147003 MM; A4870F07DC201F97 CRC64;

Query Match 84.2%; Score 32; DB 5; Length 1311;

Best Local Similarity 85.7%; Pred. No. 4.5e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEXL 8

DB 1068 LINFEXL 1074

Search completed: January 30, 2004, 07:09:11

Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:15:53 ; Search time 71 Seconds

(without alignments)
17,885 Million cell updates/sec

Title: SEQ10

Perfect score: 38

Sequence: 1 sinifexl 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 125

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%

Maximum Match 100%
Listing first 250 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	8	15	AAK57996
2	38	100.0	8	16	AAK83938
3	38	100.0	8	17	AAK89157
4	38	100.0	8	18	AAK19955
5	38	100.0	8	18	AAK14087
6	38	100.0	8	18	AAK04642
7	38	100.0	8	19	AAK80296
8	38	100.0	8	19	AAK70375
9	38	100.0	8	19	AAK68308

10	38	100.0	8	19	AAK68365	MHC binding peptid
11	38	100.0	8	19	AAK60700	Ovalbumin peptide
12	38	100.0	8	19	AAK54265	Ovalbumin peptide
13	38	100.0	8	19	AAK52838	Ovalbumin artifact
14	38	100.0	8	20	AAK42307	Ovalbumin-derived
15	38	100.0	8	20	AAK16871	OVA peptide sequen
16	38	100.0	8	20	AAK03780	Ovalbumin peptide
17	38	100.0	8	20	AAK99480	Ovalbumin-derived
18	38	100.0	8	20	AAK67585	Ovalbumin-derived
19	38	100.0	8	21	AAK13763	T-cell activation
20	38	100.0	8	21	AAK29465	Negative control p
21	38	100.0	8	21	AAK26484	tumour associated
22	38	100.0	8	21	AAK13785	OVA-derived peptid
23	38	100.0	8	21	AAK68311	Altered MHC determ
24	38	100.0	8	21	AAK59401	Ovalbumin protein
25	38	100.0	8	21	AAK52564	Marine ovalbumin M
26	38	100.0	8	21	AAK52965	Altered MHC determ
27	38	100.0	8	22	AAK77871	Ovalbumin 257-264
28	38	100.0	8	22	AAK52562	Cytotoxic T-cell e
29	38	100.0	8	22	AAK13119	Ovalbumin (OVA)-de
30	38	100.0	8	22	AAK09514	Marine ovalbumin (
31	38	100.0	8	22	AAK09514	Human ovalbumin pe
32	38	100.0	8	22	AAK63855	Amino acid sequenc
33	38	100.0	8	22	AAK66422	Chicken ovalbumin
34	38	100.0	8	22	AAK05398	Peptide released f
35	38	100.0	8	22	AAK06033	Chicken ovalbumin
36	38	100.0	8	22	AAK84316	Peptide used to pr
37	38	100.0	8	22	AAK99354	Ovalbumin cytotoxi
38	38	100.0	8	22	AAK82176	Immunodominant CTL
39	38	100.0	8	22	AAK81122	Chicken ovalbumin
40	38	100.0	8	22	AAK82065	Ovalbumin-derived
41	38	100.0	8	22	AAK92374	Miscellaneous pept
42	38	100.0	8	22	AAK89590	Ovalbumin MHC clas
43	38	100.0	8	22	AAK28959	Chicken ovalbumin
44	38	100.0	8	23	AAK26368	Ovalbumin CTL epit
45	38	100.0	8	23	AAK79933	Ovalbumin T-cell e
46	38	100.0	8	23	AAK93028	Mouse class I MHC
47	38	100.0	8	23	AAK25400	Chicken ovalbumin
48	38	100.0	8	23	AAK31661	Chicken ovalbumin
49	38	100.0	8	23	AAK31967	Chicken ovalbumin
50	38	100.0	8	23	AAK09718	Mouse MHC class I K
51	38	100.0	8	23	AAK08108	Chicken ovalbumin
52	38	100.0	8	23	AAK81273	Chicken OVA 257-26
53	38	100.0	8	23	AAK22531	Ovalbumin peptide
54	38	100.0	8	23	AAK76050	Ovalbumin, H-2kb r
55	38	100.0	8	23	AAK19945	Cytotoxic T-cell e
56	38	100.0	8	23	AAK09907	Ovalbumin peptide
57	38	100.0	8	23	AAK76942	OVA peptide (257-2
58	38	100.0	8	23	AAK76802	MHC class I-restri
59	38	100.0	8	23	AAK18665	OVA peptide fragme
60	38	100.0	8	23	AAK11866	Ovalbumin derived c
61	38	100.0	8	23	AAK11239	Immunodominant Kb-
62	38	100.0	8	23	AAK09820	Ovalbumin-derived
63	38	100.0	8	23	AAK13215	Hen egg ovalbumin
64	38	100.0	8	23	AAK13436	Chicken ovalbumin
65	38	100.0	8	23	AAK75056	Ovalbumin antigeni
66	38	100.0	8	24	AAK07743	Chicken ovalbumin
67	38	100.0	8	24	AAK08619	Ovalbumin (OVA) re
68	38	100.0	8	24	AAK57401	Synthetic bmer pep
69	38	100.0	8	24	AAK58359	Ovalbumin-derived
70	38	100.0	8	24	AAK73081	MHC class I peptid
71	38	100.0	8	24	AAK56760	Ovalbumin derived
72	38	100.0	8	24	AAK60027	Ovalbumin antigeni
73	38	100.0	8	24	AAK11029	Ovalbumin immunode
74	38	100.0	9	24	AAK84323	Peptide used to pr
75	38	100.0	9	24	AAK57402	Synthetic bmer pep
76	38	100.0	10	18	AAK04643	Ovalbumin-derived
77	38	100.0	10	18	AAK04644	Ovalbumin-derived
78	38	100.0	10	23	AAK09821	Modified ovalbumin
79	38	100.0	10	23	AAK09825	Modified ovalbumin
80	38	100.0	12	18	AAK14122	OVA protein derive
81	38	100.0	12	23	AAK09822	Modified ovalbumin
82	38	100.0	23	23	AAK09826	Modified ovalbumin

83	38	100.0	12	23	AAU09827	Modified ovalbumin
84	38	100.0	14	23	ABB76049	Peptide insert in
85	38	100.0	14	23	AAU09823	Modified ovalbumin
86	38	100.0	14	23	AAU09828	Modified ovalbumin
87	38	100.0	15	23	AAU09824	Modified ovalbumin
88	38	100.0	16	24	ABP57403	Synthetic lemer pe
89	38	100.0	19	18	AAW19957	B1P-binding domain
90	38	100.0	19	18	AAW19956	OVA-B1P-binding do
91	38	100.0	19	23	AAE13446	Chicken MHC class
92	38	100.0	19	23	AAE13447	Chicken MHC class
93	38	100.0	19	24	ABP57404	Synthetic lemer pe
94	38	100.0	24	14	AAK32294	Antigenic peptide
95	38	100.0	24	14	AAK41450	Ovalbumin-derived
96	38	100.0	24	18	AAW04645	Ovalbumin based pe
97	38	100.0	24	22	ANG65170	Myelin basic prote
98	38	100.0	24	22	ABP57439	Ovalbumin (OVA) pe
99	38	100.0	24	23	ABG31664	Synthetic 26mer pe
100	38	100.0	26	24	ABP57405	Synthetic 26mer* p
101	38	100.0	26	24	ABP57406	Synthetic MHC class
102	38	100.0	30	23	AAE13448	Synthetic 31mer pe
103	38	100.0	31	24	ABP57407	Ovalbumin-derived
104	38	100.0	35	18	AAW04646	Chicken ovalbumin
105	38	100.0	36	24	AAO26741	Amino acid sequenc
106	38	100.0	43	22	AAH84325	Amino acid sequenc
107	38	100.0	47	22	AAH84321	Amino acid sequenc
108	38	100.0	48	22	AAH84322	T55-DICE ovalbumin
109	38	100.0	49	22	AAH84953	DICE-I ovalbumin M
110	38	100.0	57	22	AAH84954	Chicken ovalbumin
111	38	100.0	100	23	AAE13458	Chicken ovalbumin
112	38	100.0	100	23	AAE13460	Chicken ovalbumin
113	38	100.0	103	23	AAE13459	Chicken ovalbumin
114	38	100.0	103	23	AAE13461	Polytope sequence.
115	38	100.0	106	17	AAH89966	Chicken ovalbumin
116	38	100.0	108	23	AAE13462	Chicken ovalbumin
117	38	100.0	111	23	AAE13463	Chicken ovalbumin
118	38	100.0	132	21	AAV52575	Amino acid sequenc
119	38	100.0	386	23	AAE13435	Chicken ovalbumin
120	38	100.0	409	22	AAH31545	Amino acid sequenc
121	38	100.0	479	22	AAE13112	Human HER300-rGM-C
122	38	100.0	541	23	AAU99725	Yeast/mouse SS-OVA
123	38	100.0	564	22	AAE13110	Human HER500 fuesio
124	38	100.0	697	22	AAE13111	Human HER500-rGM-C
125	38	100.0	948	22	AAH31611	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAR57996
ID AAR57996 strand: Protein; 8 AA.
XX
AC AAR57996;
XX
DT 25-MAR-2003 (updated)
DT 30-MAR-1995 (first entry)
XX
DE Ova257-264.
XX
XX Ova; ovalbumin; cytosol; cytolitic immune response;
KW vaccinia virus; promoter.
XX
OS Synthetic.
XX
PN WO9417816-A1.
XX
PD 18-AUG-1994.
XX
PF 27-JAN-1994; 94WO-US01183.
XX
PR 10-FEB-1993; 93US-0016066.
XX
PA (DAND) DANA FARBER CANCER INST INC.

PA (HARD) HARVARD COLLEGE.
XX Goldberg AL, Rock KL;
PI
XX WPI; 1994-279383/34.
DR
PT Method for blocking cytolytic immune responses - useful for
PT treatment of autoimmune diseases and preventing organ and graft
PT rejection
XX
PS Example 2; Page 44; 89pp; English.
XX
CC Ova257-264 was constructed by inserting a synthetic oligonucleotide
CC (AAO67342) behind the vaccinia virus p7.5 early/late promoter in psc11
CC which was modified such that the restriction sites SalI and NotI
CC were substituted for the SmaI site. The oligonucleotide consists of
CC a SalI site, Kozak's consensus sequence for efficient translation,
CC an initiation codon, nucleotides encoding the peptide given in
CC AAR57996, two stop codons, and a NotI site.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 15; Length 8;
Best local similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 STINPEKL 8
Db 1 STINPEKL 8
RESULT 2
ID AAR83938 standard; peptide; 8 AA.
XX
AC AAR83938;
XX
DT 05-JUN-1996 (first entry)
XX
DE MHC class I restricted antigenic peptide #8.
XX
KW MHC class I; antigen; MAGE; melanoma; breast cancer; bladder cancer;
KW Titermax; cytotoxic T-lymphocyte; tumour; pathogenic disease; bacteria;
KW parasite; human; animal.
XX
OS Synthetic.
XX
PN WO9528958-A1.
XX
PD 02-NOV-1995.
XX
PF 21-APR-1995; 95WO-US04975.
XX
PR 22-APR-1994; 94US-0233496.
XX
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PI Dyall R, Nikolic-Zugic J;
XX
DR WPI; 1995-382848/49.
XX
PT Cytotoxic T-cell induction by MHC class I-restricted peptide in
PT adjuvant - useful for treating tumours and bacterial or parasitic
PT pathogenic diseases
XX
PS Claim 11; Page 38; 50pp; English.
XX
CC The sequences given in AAR83931-49 are MHC class I restricted 8-12
CC amino acid antigenic peptides. This peptide represents ovalbumin
CC residues 257-264. These peptides may be administered to a subject
CC in combination with a suitable adjuvant, pref. Titermax (RTM), to
CC induce cytotoxic T-lymphocytes. This method may be used in the

CC treatment of a tumour or a pathogenic disease, esp. diseases of
 CC bacterial or parasitic origin, in humans and animals, e.g monkeys,
 CC dogs, cows, horses, etc.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 16; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8

RESULT 3

AAR89157
 ID AAR89157 standard; peptide; 8 AA.

XX AAR89157;

DT 25-MAR-2003 (updated)
 DT 03-SEP-1996 (first entry)

DE Peptide Ova8 used in an MHC stripping/reloading method.

XX Antigen; major histocompatibility complex; cell surface; stimulation;
 KW cytotoxic T lymphocyte; CTL; endogenous; exogenous; peptide; spleen;
 KW allelic restriction; peripheral blood lymphocytes; ganglion; placenta;
 KW native form; infection; tumour; autoimmune disease.

OS Synthetic.

PN WO9601891-A1.

PD 25-JAN-1996.

PF 06-JUL-1995; 95WO-FR00907.

PR 07-JUL-1994; 94FR-0008427.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PA (INSP) INST PASTEUR.

PI Langlade Demoyen P, Kourilsky P, Abastado J;

DR WPI; 1996-097621/10.

XX Cell population with high surface density of exogenous peptide bound
 PT to MHC molecules - prepd. by stripping endogenous peptide and
 PT reloading exogenous peptide(s), useful for stimulating cytotoxic
 PT lymphocytes in cases of infection, tumour and auto-immune disease
 XX
 XX Disclosure; Page 9; 37pp; French.

XX Peptides AAR89154-73 are examples of exogenous peptides which are
 CC "loaded" onto the antigen-presenting major histocompatibility complex
 CC (MHC) on the cell surface in a novel method of stimulating cytotoxic T
 CC lymphocytes. The method involves treating cells, e.g. at a pH 5 or below
 CC or pH 9 or above, to remove the endogenous peptides from the MHC,
 CC followed by recharging the complexes with specified exogenous peptides
 CC having the same allelic restriction as the MHC. The cells are pref.
 CC peripheral blood lymphocytes, spleen, ganglion or placental cells which
 CC are able to present the exogenous antigen in a native form. Recharged
 CC cells contain exogenous peptides at a higher density than native cells
 CC and are used to stimulate specific cytotoxic T lymphocytes in response to
 CC infections, tumours or autoimmune diseases. This peptide is derived from
 CC ovalbumin and belongs to the allelic restriction Kb.
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 17; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8

RESULT 4

AAW19955
 ID AAW19955 standard; peptide; 8 AA.

XX AAW19955;

DT 10-NOV-1997 (first entry)

DE Chicken OVA-peptide.

XX Vaccine; immunotherapy; heat shock protein; OVA; cancer;
 KW infectious disease.

OS Gallus sp.

PN WO9706821-A1.

PD 27-FEB-1997.

PF 16-AUG-1996; 96WO-US13363.

PR 18-AUG-1995; 95US-0002490.

PR 18-AUG-1995; 95US-0002479.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PI Hartl FU, Hoe MH, Houghton A, Mayhew M, Rothman JE;

PI Takeuchi Y;

DR WPI; 1997-165035/15.

XX Compns. for inducing immune response contg. antigen and heat shock
 PT protein - also new hybrid peptide and related nucleic acid, for
 PT treatment of infectious diseases and tumours
 XX
 XX Example 1; Page 17; 58pp; English.

XX Chicken OVA-peptide (AAW19955) is used in novel hybrid peptides,
 CC OVA-BiP (AAW19956) and BiP-OVA (AAW19957) with heat shock protein (HSP)
 CC BiP binding domain (see also AAW19951). The hybrid protein is
 CC combined in vitro with a HSP, such as hsp70, to form a complex
 CC that, when administered to a subject, induces an immune response.
 CC Vaccine compositions were prepd. by combining recombinant mouse
 CC hsp70, recombinant human hsp40 and Ova-peptide. Combinations of
 CC antigen with hsp70 or a mixture of hsp70 and hsp40 were effective
 CC to produce a cytotoxic T lymphocyte response.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 18; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8

RESULT 5

AAW14087
 ID AAW14087 standard; peptide; 8 AA.

XX AAW14087;

DT 20-OCT-1997 (first entry)

```

XX MHC class I molecule binding OVA peptide.
DE
XX Major histocompatibility complex; MHC; target; binding; tumour;
KW cancer; neoplasia; USTR; EL-4; identification; detection; screening;
KW tissue typing; Bcr-abl.
XX
XX Mus sp.
OS
XX WO9641188-A1.
PN
XX 19-DEC-1996.
PD
XX 07-JUN-1996; 96WO-US09680.
PF
XX 07-JUN-1995; 95US-0485610.
PR
XX (UNITW ) UNIV WASHINGTON.
PA
XX Cheever MA, Chen W;
PI
XX WPI; 1997-108657/10.
DR
XX Identifying major histocompatibility complex class I binding mols. -
PT using peptide(s) having a core of 7-14 amino acids with extra amino
PT acids and a reporter gp. at the N- or C-terminus, useful for tissue
PT typing
XX
XX Example 2, Page 19; 41pp; English.
PS
XX AAW4087-91 are peptides derived from ISTR and EL-4 tumours of Balb/c
CC mice. The peptides were tested for MHC specificity to find MHC I
CC specific peptides. These peptides are useful for tissue typing or for
CC screening for molecules that interact with MHC class I molecules. MHC
CC class I molecules can be identified using the peptides and also the
CC peptides are useful in vaccines against disease and infection e.g.
CC caused by viruses, bacteria or tumours.
CC
SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 18; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEKL 8
DB 1 STINFEKL 8

RESULT 6
AAW04642
ID AAW04642 standard; peptide; 8 AA.
XX
AC AAW04642;
XX
DT 01-AUG-1997 (first entry)
XX
DE Ovalbumin-derived activated CD8+ T cells epitope OVA8.
XX
KW Macrophage; artificial antigen presenting cell; APC; cancer;
KW tumours; neoplasia; viral infection; retroviral infection;
KW autoimmune.
XX
OS Synthetic.
XX
XX WO9637107-A1.
PN
XX 28-NOV-1996.
PD
XX 22-MAY-1996; 96WO-US07436.
PF
XX 23-MAY-1995; 95US-0447761.
PR
XX
XX

```

```

PA (SCRI ) SCRIPPS RES INST.
XX
XX DeBrujn MLH, Jackson MR, Peterson PA;
PI
XX WPI; 1997-020850/02.
DR
XX
XX Prodn. of activated CD8+ T cells directed to specific antigen - can
PT specifically kill target cells useful to treat, e.g. cancer
PT
XX
XX Example 1; Page 26; 84pp; English.
PS
XX
XX The method for the production of activated CD8+ T cells specifically
CC directed towards a particular antigen involves affixing peptides
CC corresponding to the particular antigen to an artificial support;
CC contacting macrophages with the affixed peptides for a time sufficient
CC for the peptides to be engulfed, and at least a portion of the peptides
CC to be presented on the surface of the macrophage; and contacting
CC unprimed CD8+ T cells with the peptide presenting macrophages for a
CC time sufficient to activate the unprimed CD8+ T cells. The present
CC sequence represents a peptide designated OVA8 which corresponds to
CC ovalbumin, a Kb-restricted peptide antigen. This is the optimal
CC peptide. Small extensions to the optimal peptide affect the affinity
CC of the peptide for soluble class I molecules in vitro e.g. the addition
CC of two amino acids to the amino-terminus lowers the affinity to Kb by
CC 76-fold compared to the optimal peptide; addition of two amino acids to
CC the carboxy-terminus lowers the affinity by 4-fold. The method,
CC macrophages and artificial antigen presenting cell, having a peptide
CC corresponding to the particular antigen present on its surface and at
CC least a portion of an artificial support in its interior, can be used to
CC treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
CC infection or autoimmune or autoimmune-type conditions) in patients via
CC the specific killing of target cells.
CC
SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 18; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STINFEKL 8
DB 1 STINFEKL 8

RESULT 7
AAW80296
ID AAW80296 standard; Peptide; 8 AA.
XX
AC AAW80296;
XX
DT 08-JAN-1999 (first entry)
XX
DE Amino acids 257-264 of chicken ovalbumin used as an antigen.
XX
KW Antisense oligonucleotide; antigen processing protein; TAP;
KW transporter; proteasome; antigen-presenting cell; cancer; infection;
KW cytotoxic T cell; chicken ovalbumin.
XX
OS Synthetic.
OS Gallus sp.
XX
XX US5831068-A.
PN
XX 03-NOV-1998.
PD
XX 20-AUG-1996; 96US-0700035.
PF
XX 20-AUG-1996; 96US-0700035.
PR
XX 21-AUG-1995; 95US-0517373.
PR
XX (UYDU-) UNIV DUKE.
PA
XX
XX Gilboa E, Nair SK;
PI

```

XX DR WPI; 1998-609331/51.
 XX PT Increasing the presentation of a peptide on a mammalian cell for
 XX production of antigen-presenting cells and stimulation of immune
 PT response - by contacting cells with antigen after inactivating the
 PT protein transporter associated with antigen processing or proteasome
 XX
 PS Disclosure; Column 12; 27pp; English.
 XX
 CC AAM80296-99 represent peptide antigens used in the course of the
 CC invention. The specification describes a method for increasing the
 CC presentation of a peptide (antigen) on a mammalian cell. The method
 CC comprises inhibiting the activity of a transporter associated with
 CC TAP or proteasome in the cell in vitro before contacting the cell
 CC with the peptide. Antigen-presenting cells produced as above can be used
 CC to stimulate an immune response in vitro or in vivo e.g. to treat or
 CC prevent cancer or infection with a pathogen, e.g. a bacterium or virus.
 CC Cytotoxic T cells produced as above can also be used for therapy.
 CC
 SQ Sequence 8 AA;
 OY Query Match 100.0%; Score 38; DB 19; Length 8;
 AC Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SIINPEKL 8
 DB 1 SIINPEKL 8
 ID AAM70375 standard; Protein; 8 AA.
 AC AAM70375;
 XX
 DT 18-NOV-1998 (first entry)
 XX
 DE Ovalbumin peptide used in the method of the invention.
 XX
 KW Ovalbumin; hep70; heat shock protein; vaccine; tumour therapy.
 XX
 OS Synthetic.
 OS Gallus sp.
 XX
 PN WO9835705-A1.
 XX
 PD 20-AUG-1998.
 XX
 PF 18-FEB-1998; 98WO-US03033.
 XX
 FR 25-NOV-1997; 97US-0066288.
 PR 18-FEB-1997; 97US-0038059.
 XX
 PA (MHED) WHITEHEAD INST BIOMEDICAL RES.
 XX
 PI Young RA;
 XX
 DR WPI; 1998-456872/39.
 XX
 PT Use of heat shock protein - for delivery of moiety into cells,
 PT useful for vaccination against tumours
 XX
 PS Disclosure; Fig 1A; 45pp; English.
 XX
 CC The present sequence represents residues 258-276 of the ovalbumin
 CC protein. The ovalbumin peptide was used in the method of the
 CC invention. The invention provides a method for delivering a moiety
 CC (e.g. ovalbumin protein) of interest into a cell which involves
 CC contacting the cell with a complex comprising the moiety of interest
 CC covalently linked to a heat shock protein (e.g. hep70). The method
 CC is claimed to be useful for providing the efficient delivery into cells

CC of moieties which are not normally able to enter cells or which enter
 CC cells only to a limited extent. The method can be useful for delivering
 CC moieties such as proteins, peptides, lipids, glycoproteins, small
 CC organic molecules and other molecules, particularly chemicals and other
 CC molecules which are useful therapeutically or diagnostically. The
 CC method is claimed to be useful when applied to vaccination regimes,
 CC e.g. for tumour therapy.
 XX
 SQ Sequence 8 AA;
 OY Query Match 100.0%; Score 38; DB 19; Length 8;
 AC Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SIINPEKL 8
 DB 1 SIINPEKL 8
 ID AAM68308 standard; peptide; 8 AA.
 AC AAM68308;
 XX
 DT 25-MAR-2003 (updated)
 DT 14-OCT-1998 (first entry)
 XX
 DE MHC binding peptide Ova8.
 XX
 KW Antigen; major histocompatibility complex; MHC; lymphocyte; detection;
 KW immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;
 KW viral infection.
 XX
 OS Synthetic.
 OS WO9744667-A2.
 XX
 PN 27-NOV-1997.
 XX
 PD 21-MAY-1997; 97WO-FR00892.
 XX
 PF 21-MAY-1996; 96US-0651925.
 PR
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (INSP) INST PASTEUR.
 XX
 PI Langlademeyen P, Lone Y, Kourilsky P, Abastado J;
 XX
 DR WPI; 1998-018653/02.
 XX
 XX Detection, purification and elimination of antigen-specific
 PT lymphocytes - for producing cytotoxic T cells for immuno-therapy of
 PT cancers and viral infection
 XX
 PS Disclosure; Page 24; 222pp; French.
 XX
 CC Peptides AAM68301-W68384 are examples of antigens (Ag) which can be
 CC loaded onto recombinantly produced major histocompatibility complex
 CC (MHC) molecules in a method of detecting antigen-specific lymphocytes.
 CC The MHC-antigen complex is then immobilised on a solid support and a
 CC sample containing cells recognising the MHC-Ag complex may be isolated.
 CC A similar method is used to isolate, purify or eliminate Ag-specific
 CC T-cells or to produce Ag-specific cytotoxic T-cells (CTC). The method
 CC is also used to detect and quantify tumour-specific T-cells and to
 CC generate CTC for specific killing of tumour cells (solid tumours,
 CC leukaemia or lymphoma) by injection into a human or animal, but also
 CC for treating viral infections.
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX
 SQ Sequence 8 AA;
 OY Query Match 100.0%; Score 38; DB 19; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
Db 1 SIINFEKL 8

RESULT 10

AAW68365
ID AAW68365 standard; peptide; 8 AA.

AC AAW68365;

XX 25-MAR-2003 (updated)
DT 14-OCT-1998 (first entry)

DE MHC binding peptide from ovalbumin.

XX Antigen: major histocompatibility complex; MHC; lymphocyte; detection;
KW immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;
KM viral infection.

XX Synthetic.

XX MO9744667-A2.

XX 27-NOV-1997.

XX 21-MAY-1997; 97WO-FR00892.

XX 21-MAY-1996; 96US-0651925.

PA (INRM) INSENM INST NAT SANTE & RECH MEDICALE.
PA (INSP) INST PASTEUR.

PI Langlademeyen P, Lone Y, Kourilsky P, Abateado J;

DR WPI; 1998-018653/02.

PT Detection, purification and elimination of antigen-specific
PT lymphocytes - for producing cytotoxic T cells for immuno-therapy of
PT cancers and viral infection
XX
XX
PS Disclosure; Page 29; 222pp; French.

CC Peptides AAW68301-W68384 are examples of antigens (Ag) which can be
CC loaded onto recombinantly produced major histocompatibility complex
CC (MHC) molecules in a method of detecting antigen-specific lymphocytes.
CC The MHC-antigen complex is then immobilised on a solid support and a
CC sample containing cells recognising the MHC-Ag complex may be isolated.
CC This peptide is derived from amino acids 258-276 of ovalbumin. A
CC similar method is used to isolate, purify or eliminate Ag-specific
CC T-cells or to produce Ag-specific cytotoxic T-cells (CTC). The method is
CC also used to detect and quantify tumour-specific T-cells and to generate
CC CTC for specific killing of tumour cells (solid tumours, leukaemia or
CC lymphoma) by injection into a human or animal, but also for treating
CC viral infections.
CC (Updated on 25-MAR-2003 to correct PI field.)
CC
XX
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 19; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
Db 1 SIINFEKL 8

RESULT 11
AAW60700

ID AAW60700 standard; peptide; 8 AA.

XX AAW60700;

XX 22-SEP-1998 (first entry)

DE Ovalbumin peptide Tc1 peptide epitope (residues 257-264).

XX Immunisation; target antigen; epitope; inoculation; infant mammal;
KW viral antigen; depressed humoral response; respiratory syncytial virus;
KW rotavirus; measles virus; human immunodeficiency virus; hepatitis virus;
KW herpes simplex virus; influenza virus; Streptococcus pneumoniae;
KW Hemophilus influenzae; Neisseria meningitidis; Staphylococcus aureus;
KW protozoan antigen; malaria.

XX Unidentified.

XX MO9822145-A1.

XX 28-MAY-1998.

XX 21-NOV-1997; 97WO-US21687.

XX 22-NOV-1996; 96US-0755034.

PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.

PI Bona C, Bot A;

DR WPI; 1998-312182/27.

PT Immunisation of infant mammals - by inoculating the mammal with a
PT nucleic acid encoding a relevant epitope of a target antigen
XX
XX
PS Disclosure; Page 10; 83pp; English.

XX Sequence shown in AAW60683 to AAW60700 are epitope sequences of various
CC viral antigens used to exemplify the method of invention of immunising
CC an infant mammal against a target antigen. The method comprises
CC inoculating the mammal with a nucleic acid encoding a relevant epitope
CC of a target antigen in a carrier, such that the relevant epitope is
CC expressed in the infant mammal. The genetic immunisation of infant
CC mammals can give rise to effective cellular (including the induction of
CC cytotoxic T lymphocytes) and humoral immune responses against the target
CC antigen. The methods are particularly used for treating infants with
CC depressed humoral responses, that have high-zone tolerances against the
CC target antigens or have a Th2 biased immune response. The target antigen
CC may be a viral antigen, e.g. a respiratory syncytial virus antigen, a
CC rotavirus antigen, a measles virus antigen, a human immunodeficiency
CC virus antigen, a hepatitis virus antigen, a hepatitis B virus antigen, a
CC herpes simplex virus antigen or an influenza virus antigen, a bacterial
CC antigen e.g. Streptococcus pneumoniae antigen, Hemophilus influenzae
CC antigen, Neisseria meningitidis antigen, Staphylococcus aureus antigen
CC or a protozoan antigen such as a malaria antigen.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 19; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
Db 1 SIINFEKL 8

RESULT 12

AAW54265
ID AAW54265 standard; peptide; 8 AA.

XX AAW54265;

DT 30-JUL-1998 (first entry)

XX Ovalbumin peptide OVA.
 DE Pep-MHC complex; cytotoxic; T cell; cancer; ovarian; brain;
 KW Major histocompatibility complex.
 XX
 XX Synthetic.
 OS
 XX MO9807441-A1.
 PN
 XX 26-FEB-1998.
 PD
 XX 22-AUG-1997; 97WO-US14814.
 PF
 XX 23-AUG-1996; 96US-0023437.
 PR
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PA (UNIT) UNIV ILLINOIS FOUNO.
 XX
 XX Eisen HN, Kranz DM;
 PI
 XX WPI; 1998-168897/15.
 DR
 XX Composition for targeting an allo-reactive response to specific
 PT cells - comprises MHC-peptide complex bound to specific targeting
 PT agent, the complex having at least one allogenic component,
 PT specifically for treating ovarian and brain cancers
 XX
 XX Examples; Page 26; 45pp; English.
 PS
 XX Peptide OVA derived from ovalbumin was used in the production of soluble
 CC pep-MHC complexes. The pep-MHC (peptide-major histocompatibility
 CC complex) complexes, including both MHC heavy and light chains, bound to a
 CC specific targeting molecule can be used to target a cytotoxic T cell
 CC response to specific cells. This is particularly useful for targeting
 CC cancer cells, specifically of the ovary or brain. The method can also
 CC be used to eliminate an entire cell type for example during bone marrow
 CC therapy.
 CC
 SQ Sequence 8 AA;
 QY
 DB 1 SIINFEKL 8
 1 SIINFEKL 8
 Query Match 100.0%; Score 38; DB 19; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PA (UYPI-) UNIV PITTSBURGH.
 XX
 XX Falo LD, Rock KL;
 PI
 XX WPI; 1998-086733/08.
 DR
 XX Inducing anti-tumour cytotoxic T lymphocytes - by cross-priming
 PT with artificial antigen, then immunisation with tumour cells
 PT modified in vitro to express the same antigen, does not require
 PT characterisation of tumour-specific antigens
 XX
 XX Example; Page 21; 44pp; English.
 PS
 XX Ovalbumin can be used as an artificial target antigen (ATA) to
 CC promote a cytotoxic T lymphocyte mediated response in mammals. Tumour
 CC cells from the host can be engineered to include ATA and therefore
 CC induce anti-tumour cytotoxic T lymphocytes. This method can be used for
 CC the treatment and prevention of a wide range of tumours even when
 CC the tumour is inaccessible or where metastases are being targeted.
 CC
 SQ Sequence 8 AA;
 QY
 DB 1 SIINFEKL 8
 1 SIINFEKL 8
 Query Match 100.0%; Score 38; DB 19; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
 ID AAY42307
 ID AAY42307 standard; peptide; 8 AA.
 AC AAY42307;
 XX
 XX 06-DEC-1999 (first entry)
 DT
 DE Ovalbumin-derived peptide antigen.
 XX
 XX Immunity; human leukocyte antigen; HLA; MHC; antigen;
 KW major histocompatibility complex; presentation; solubility;
 KW dendritic cell.
 XX
 OS Synthetic.
 XX
 XX WO9947646-A1.
 PN
 XX 23-SEP-1999.
 PD
 XX 19-MAR-1999; 99WO-US06627.
 PF
 XX 20-MAR-1998; 98US-0078832.
 PR
 XX (LUDW-) LUDWIG INST CANCER RES.
 PA (BIOP-) BIOPOLIO SCRL.
 XX
 XX Rescigno M, Girolomoni G, Corinti S, Rlociardi-Caestagnoli P;
 PI
 XX WPI; 1999-571834/48.
 DR
 XX Preparation of dendritic cells which present antigens, used for
 PT stimulating an immune response by immunocompetent cells
 PT
 XX Example 6; Page 13; 43pp; English.
 PS
 XX This sequence represents an ovalbumin-derived peptide antigen which
 CC is presented on major histocompatibility complex (MHC) Class I molecules
 CC of dendritic cells via the use of a novel process to improve soluble
 CC protein antigen presentation. This process uses dendritic cells which
 CC have internalised bacterial cells recombinantly expressing ovalbumin. The
 CC dendritic cells which have internalised the bacteria can be used to

CC generate an immune response which is stronger than the response which
 CC would be generated when soluble antigen alone is used. The dendritic
 CC cells which result from the internalisation are characterised by MHC
 CC Class I and Class II molecules which have much longer half lives and
 CC greater stability than comparable dendritic cells which have not
 CC internalised such bacteria. The dendritic cells can be contacted with
 CC immunocompetent cells for stimulating an immune response. The dendritic
 CC cells can also be used to stimulate maturation of an immature dendritic
 CC cell.

XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 15

AAV16871 AAV16871 standard; peptide; 8 AA.

XX AAV16871;

DT 20-JUL-1999 (first entry)

XX OVA peptide sequence.

XX Conjugate peptide; heat shock protein; hsp; phage display library; virus;
 KM surface protein; tethering peptide; chaperone process; cytokine; cancer;
 KM neoplastic disease; infectious disease; bacterium; immune system; fungus;
 KM acquired immune deficiency; autoimmune disease.

XX Synthetic.

XX MO9922761-A1.

XX 14-MAY-1999.

XX 22-OCT-1998; 98WO-US22335.

XX 31-OCT-1997; 97US-0961707.

XX (SLOAN KETTERING INST CANCER RES.

XX Hartl U, Hoe MH, Houghton A, Mayhew M, Moroi Y;
 PI Querfeldt O, Rothman JF;

XX WPI; 1999-313177/26.

XX Identifying peptides which bind heat shock proteins

XX Examples; Page 49; 155pp; English.

XX The invention relates to conjugate peptides engineered to noncovalently
 CC bind to heat shock proteins (hsp). A method of identifying a hsp binding
 CC peptide comprises (a) contacting a phage display library having
 CC bacteriophage expressing, in a surface protein, inserted peptides with a
 CC hsp target, and bound to a benzocoumarin anisoyl antibiotic (BMA), in a
 CC physiologic binding buffer; (b) isolating a phage binding to the hsp
 CC target; and (c) identifying the inserted peptide expressed. The peptides
 CC which bind to a hsp can be used as tethering peptides for a hsp which may
 CC serve as an accessory in a chaperone process and/or may comprise a
 CC cytokine. They can also be coupled to antigens to induce an immune
 CC response. Such compositions can be used for treating neoplastic disease,
 CC e.g. cancers, infectious diseases, e.g. diseases caused by a bacterium,
 CC virus, protozoan, mycoplasma, fungus, yeast, parasite or prion, or a
 CC disease of the immune system, e.g. acquired immune deficiencies or
 XX autoimmune diseases.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 16

AAV03780 AAV03780 standard; peptide; 8 AA.

XX AAV03780;

DT 23-JUN-1999 (first entry)

XX Ovalbumin peptide fragment (residues 257-264).

XX Dendritic cell-derived factor; proliferation; interferon gamma; IFNgamma;
 KM T cell; granulocyte macrophage colony-stimulating factor; GM-CSF;
 KM T cell stimulatory factor; lymphocyte; interleukin-2; lipopolysaccharide;
 KM autoimmune response; inflammatory; ovalbumin.

XX Synthetic.

XX Gallus gallus.

XX MO9918909-A2.

XX 22-APR-1999.

XX 14-OCT-1998; 98WO-US21614.

XX 14-OCT-1997; 97US-0062405.

XX (LUDWIG INST CANCER RES.

XX Dunn A, Marino MW, Noguchi Y, Old LJ, Wada H;

XX WPI; 1999-277418/23.

XX Dendritic cell and T cell derived factors for regulation of T cell
 PT proliferation and interferon gamma production

XX Examples; Page 37; 68pp; English.

XX The invention relates to a dendritic cell-derived factor that restores
 CC proliferation and interferon gamma (IFNgamma) production to T cells from
 CC granulocyte macrophage colony-stimulating factor (GM-CSF). The dendritic
 CC cell-derived factor which is a T cell stimulatory factor modulates the
 CC effect of GM-CSF on production of IFNgamma by lymphocytes and the
 CC response of GM-CSF -/- T cells to interleukin-2, and corrects the
 CC lipopolysaccharide-induced defect in IFNgamma production. The dendritic
 CC cell-derived factor is used in vivo to increase proliferation of T cells
 CC and/or IFNgamma production, e.g. during immunisation to increase the
 CC response to an antigen. Agents with T cell derived factor activity are
 CC used to decrease production of IFNgamma, e.g. to reduce autoimmune
 CC responses and inflammatory reactions. Agents that bind to the factors,
 CC e.g. antisense sequences, are used to treat excessive/inadequate T cell
 CC proliferation and IFNgamma production. The dendritic cell-derived factor
 CC and T cell factor (or their fragments) can also be used to raise
 CC antibodies and as components in immunoassay and diagnostic systems, also
 CC for in vitro studies. The present sequence represents an ovalbumin
 CC peptide fragment used in the course of the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 KW 1 SIINFEXL 8

RESULT 17

AAW99480 standard; peptide; 8 AA.

AAW99480;

08-JUN-1999 (first entry)

Ovalbumin-derived peptide OVA.

Matrix protein; conjugate; mutant; major histocompatibility complex; MHC; class I molecule; beta-2-microglobulin; stimulation; immunity; tumour.

Synthetic.

WO9911775-A1.

11-MAR-1999.

20-AUG-1998; 98WO-US17308.

29-AUG-1997; 97US-0920413.

(GENO) GEN HOSPITAL CORP.
 (HARD) HARVARD COLLEGE.

Garboczi DN, Walker J;

WPI; 1999-205182/17.

Method for conjugating a mutant major histocompatibility complex class I molecule and a compound - useful for stimulating immunity in an individual, and eradicating undesired cells, especially tumours

Example III-4; Page 22; 37pp; English.

The invention relates to the preparation of a conjugate of a mutant major histocompatibility complex (MHC) class I molecule (containing a Tyre7Cys amino acid substitution in the beta2-microglobulin subunit) and a compound. This sequence corresponds to the ovalbumin-derived peptide OVA. The peptide is used as a control to prepare a hybrid peptide-MHC class I tetramer in which a mutant human beta2-microglobulin subunit binds a mouse MHC class I molecule. The conjugates are useful for stimulating immunity in an individual, and eradicating undesired cells (e.g. tumours).

Sequence 8 AA;

Query Match 100.0%; Score 38; DB 20; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 18

AAW67585 standard; peptide; 8 AA.

AAW67585;

02-MAR-1999 (first entry)

T-cell activation peptide #3.

Activated T helper cell; CD4+; cytotoxic T cell; CD8+; liposome; epitope;

peripheral blood lymphocyte; antigen-presenting cell; APC; virus; tumour; bacterium; parasite; cytokine; vaccine; cancer; malaria; HIV; hepatitis; tuberculosis.

Synthetic.

WO9850527-A1.

12-NOV-1998.

07-MAY-1998; 98WO-US09288.

08-MAY-1997; 97US-0045949.

(BIOM-) BIOMIRA INC.

Agarwal B, Krantz MJ, Longenecker BM, Reddish MA;

WPI; 1999-034715/03.

Method of activation of T cells - by exposure to antigen-presenting cells loaded with antigen in liposome, used for, e.g. treating cancer and microbial infections

Disclosure; Page 6; 75pp; English.

Peptides AAW67583-W67611 are used to produce activated T helper (CD4+) and cytotoxic (CD8+) T cells. The activated T cells are produced by treating peripheral blood lymphocytes with liposome-encapsulated peptide antigen to generate Ag-loaded antigen-presenting cells (APC), contacting native or anergic T-cells with these APC, and isolating the resulting activated T-cells. The cells are specific for a particular antigen, particularly one derived from a tumour, but also those from viruses, bacteria and other parasites. It can also be used to identify antigens and epitopes able to generate an Ag-specific T-cell response (by assessing proliferation and cytokine release). Also the Ag-loaded APC can be used as cellular vaccines for treating cancer (claimed) or other diseases (e.g. malaria, human immune deficiency virus infection, hepatitis, tuberculosis). The activated T-cells can be used to treat the same conditions by adoptive T-cell transfer therapy.

Sequence 8 AA;

Query Match 100.0%; Score 38; DB 20; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 19

AAAB13763

AAAB13763 standard; peptide; 8 AA.

AAAB13763;

02-FEB-2001 (first entry)

Peptide fragment from ovalbumin OVA.

T-cell; immune response; antigen; epitope; B7 family molecule;

Leukocyte function-associated antigen-3; LFA-3;

Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;

colon polyp; Crohn's disease; ulcerative colitis; breast lesion;

tumour; ovalbumin.

Unidentified.

WO200034494-A1.

15-JUN-2000.

```

XX 12-NOV-1999; 99WO-US26866.
XX
XX 09-DEC-1998; 98US-011582.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (THER-) THERION BIOLOGICS CORP.
XX
XX Schlom J, Hodge J, Panicali D;
XX
XX WPI, 2000-431307/37.
XX
XX Novel recombinant vector useful as immunogens and vaccines for
XX stimulating and enhancing immunological responses to target cells and
XX antigens expresses multiple co-stimulatory molecules such as B7-1,
XX LFA-3, ICAM-1
XX
XX Example 31; Page 80; 186pp; English.
XX
XX Costimulatory molecules have important roles in T-cell activation and
XX therefore the immune response. The present invention relates to
XX recombinant vectors which comprise of foreign nucleic acid sequences
XX encoding at least three costimulatory molecules: a B7 family molecule,
XX leukocyte function-associated antigen-3 (LFA-3, human CD58) and
XX intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
XX gene encoding a target antigen or immunological epitope. The present
XX sequence is one such target antigen used in the present invention. The
XX present sequence is a tumour-associated antigen. The vector of the
XX present invention would be useful for providing an enhanced immune
XX response to the present target antigen. The vector of the present
XX invention may therefore be useful in immunotherapy for treating or
XX preventing diseases caused by viruses, bacteria, protozoans, parasites,
XX prealignant cells and tumour cells. The recombinant vector can be used
XX to treat or prevent preneoplastic or hyperplastic states such as colon
XX polyps, Crohn's disease, ulcerative colitis and breast lesions.
XX
XX Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 21; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 STINFEKL 8
XX |||||
XX DB 1 STINFEKL 8
XX
XX RESULT 20
XX AAB29465
XX ID AAB29465 standard; peptide; 8 AA.
XX
XX AC AAB29465;
XX
XX DT 09-FEB-2001 (first entry)
XX
XX DE Negative control peptide used in a cytotoxic T cell assay.
XX
XX KW Telomerase antigen variant; HLA-A2-binding; class I MHF;
XX human leukocyte antigen; major histocompatibility complex;
XX cytotoxic T-cell response; antigen-presenting cell; APC;
XX telomerase-expressing cell; cancer; anticancer vaccine.
XX
XX OS Synthetic.
XX
XX PN WO200061766-A2.
XX
XX PD 19-OCT-2000.
XX
XX PF 07-APR-2000; 2000WO-IB00610.
XX
XX PR 09-APR-1999; 99US-0128539.
XX
XX PA (BIOM-) BIOMIRA INC.

```

```

XX Agrawal B, Longenecker BM;
XX
XX WPI, 2000-679493/66.
XX
XX New telomerase-specific T-cell antigens useful for generating T-cell
XX responses against telomerases and for producing vaccines for treating
XX or preventing cancer by in vivo or ex vivo techniques
XX
XX Example 1; Page 23; 34pp; English.
XX
XX The invention relates to a human telomerase peptide antigen (AAB29461)
XX which binds to a class I HLA (human leukocyte antigen, MHC, major
XX histocompatibility complex), and to conservatively substituted variants
XX thereof. The invention also relates to a vaccine comprising a telomerase
XX antigen or antigen variant, a nucleotide encoding a telomerase antigen
XX or variant, and a method of producing telomerase-primed antigen-
XX presenting cell (APC) comprising contacting an APC with a composition
XX containing a telomerase antigen or variant. The telomerase antigens or
XX vaccine compositions are useful for inducing a cytotoxic T-cell immune
XX response against telomerase and hence against telomerase-expressing
XX cells (i.e., cancer cells. Additionally, the telomerase antigen-primed
XX APC may be administered with interleukin-2 for cancer treatment or
XX prevention. The present sequence represents a peptide used in the
XX exemplification of the invention in an assay of the cytotoxic activity of
XX T-cells.
XX
XX Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 21; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 STINFEKL 8
XX |||||
XX DB 1 STINFEKL 8
XX
XX RESULT 21
XX AAB26484
XX ID AAB26484 standard; peptide; 8 AA.
XX
XX AC AAB26484;
XX
XX DT 16-JAN-2001 (first entry)
XX
XX DE Tumour associated OVA peptide.
XX
XX KW Immune response; vaccine; cancer; infection; tumour; OVA.
XX
XX OS Unidentified.
XX
XX PN WO200050080-A1.
XX
XX PD 31-AUG-2000.
XX
XX PF 23-FEB-2000; 2000WO-US04565.
XX
XX PR 26-FEB-1999; 99US-0261473.
XX
XX PA (UYDU-) UNIV DUKE.
XX
XX PI Gilboa E, Nair SK, Nicchitta CV;
XX
XX DR WPI, 2000-558368/51.
XX
XX PT Eliciting immune response in vertebrate for prevention and treatment of
XX cancer and infectious diseases involves administering purified complex
XX comprising calreticulin bound to an antigenic molecule
XX
XX PS Disclosure, Page 7; 82pp; English.
XX
XX The present invention relates to a method of eliciting an immune

```

CC response by administering a composition of a purified complex of
 CC calreticulin bound to an antigenic molecule. The present invention is
 CC useful for prevention and treatment of cancer and infectious disease
 CC in a vertebrate especially of humans. The present sequence is the
 CC tumour antigenic peptide OVA which was used in the invention.

XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 DB 1 SIINFEKL 8

RESULT 22

AAB13785
 ID AAB13785 standard; peptide; 8 AA.

XX
 AC AAB13785;

XX
 DT 10-NOV-2000 (first entry)

XX
 DE OVA-derived peptide.

XX
 KW Chicken; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;
 KW major histocompatibility complex class I; MHC class I; antigen; tumour;
 KW prostate; breast; multiple myeloma; OVA peptide.

XX
 OS Gallus domesticus.

XX
 PN WO200035949-A1.

XX
 PD 22-JUN-2000.

XX
 PF 14-DEC-1999; 99WO-US29724.

XX
 PR 14-DEC-1998; 98US-0112324.

XX
 PA (DEND-) DENDREON CORP.

XX
 PI Laus R, Hakim I, Vidovic D;

XX
 DR WPI; 2000-442365/38.

XX
 PT Antigens modified by the covalent addition of a peptide that
 PT facilitates entry into antigen presenting cells, useful for producing
 PT compositions for immunizing against tumors and pathogens -

XX
 PS Example 1; Page 10; 34pp; English.

XX
 CC The present invention relates to compositions of modified soluble protein
 CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
 CC response i.e. a major histocompatibility complex (MHC) class I molecule
 CC response. The protein antigen is modified by the covalent addition of a
 CC peptide sequence which facilitates entry of the antigen into antigen
 CC presenting cells (APCs). The present sequence is a peptide derived from
 CC the chicken antigen OVA. This peptide was used to prepare the modified
 CC antigen. The modified antigen composition may be used for immunising
 CC against, or treating a tumour e.g. prostate and breast carcinoma or
 CC multiple myeloma, or pathogen in mammals.

XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 DB 1 SIINFEKL 8

RESULT 23

AAV68311
 ID AAV68311 standard; Peptide; 8 AA.

XX
 AC AAV68311;

XX
 DT 13-APR-2000 (first entry)

XX
 DE Altered MHC determinant binding peptide SEQ ID NO:143.

XX
 KW MHC class I; major histocompatibility complex; microglobulin; antigen;
 KW immune response; immunisation; AIDS; multiple sclerosis; toxic shock;
 KW cancer; lupus erythematosus; snake bite; cytotoxic; antiviral;
 KW immunomodulatory; dermatological; immunosuppressive; antiinflammatory;
 KW neuroprotective.

XX
 OS unidentified.

XX
 PN US6011146-A.

XX
 PD 04-JAN-2000.

XX
 PF 07-JUN-1995; 95US-0481985.

XX
 PR 15-NOV-1991; 91US-0792473.

XX
 PR 05-DEC-1991; 91US-0801818.

XX
 PA (INSP) INST PASTEUR.

XX
 PA (INRM) INST NAT SANTE & RECH MEDICALE.

XX
 PI Kourilsky P, Morteiz E, Abastado J;

XX
 DR WPI; 2000-125951/11.

XX
 PT New recombinant DNA encoding covalently linked form of major
 PT histocompatibility complex Class I determinant, used for immune system
 PT stimulation, e.g. for treating cancer -

XX
 PS Disclosure; Column 12; 88pp; English.

XX
 CC The present invention describes a recombinant DNA molecule (I)
 CC containing a sequence (Ia) that encodes an altered MHC (major
 CC histocompatibility complex) Class I determinant (II) comprises a
 CC polypeptide with alpha1, alpha2, alpha3 and beta2-microglobulin
 CC domains, in which alpha3 and beta2 are covalently linked, thorough C-
 CC and N-terminal respectively, via a nucleotide spacer sequence encoding a
 CC polypeptide. (II) includes an antigen-binding site and when (II) and
 CC the antigen are associated they are recognized by a mammalian T cell
 CC receptor (TCR). (I) are used to produce (II) which are used to study
 CC functional interactions between the various MHC domains. They can also
 CC be used to modulate (in vivo or in vitro) the immune system by inducing
 CC an effector response (cytotoxicity, antibody synthesis, phagocytosis)
 CC of immune system cells, typically for treating, or immunising against;
 CC cancer, acquired immune deficiency syndrome, lupus erythematosus,
 CC multiple sclerosis, toxic shock and snake bite, but also for selective
 CC destruction of autoreactive cells, diagnostically to assay T cell
 CC receptors and to raise specific antibodies (useful for diagnosis,
 CC therapy, studying MHC-associated cellular processes and for affinity
 CC purification). AA25758 and AAV68186 to AAV68316 are sequences used in
 CC the exemplification of the present invention.

XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 DB 1 SIINFEKL 8

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINPEKL 8
1 SIINPEKL 8

RESULT 26
AAV52965
ID AAV52965 standard; Peptide; 8 AA.

AC AAV52965;

DT 14-FEB-2000 (first entry)

DE Altered MHC determinant binding peptide SEQ ID NO:143.

XX Major histocompatibility complex; MHC class I; MHC class II; antigen;
KW immune response; diagnosis; antibody; immunisation; autoimmune disease;
KW acquired immune deficiency syndrome; AIDS; cytotoxic; dermatological;
KW anti-inflammatory; neuroprotective; immunosuppressive; antithyroid;
KW vaccine; lupus erythematosus; multiple sclerosis; thyroiditis;
KW toxic shock; tumour; snakebite.

OS Synthetic.

XX US5976551-A.

PD 02-NOV-1999.

PF 07-JUN-1995; 95US-0484905.

PR 05-DEC-1991; 91US-0801818.
PR 15-NOV-1991; 91US-0792473.

XX (INSP) INST PASTEUR.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Kourilsky P, Mottiez E, Abastado J;

DR WPI; 2000-037081/03.

PT Composition containing an antigen and altered major histocompatibility
PT Class II determinant, used to immunize against autoimmune diseases,
PI e.g. acquired immune deficiency syndrome

PS Claim 8; Column 13; 96pp; English.

XX The present invention describes a composition capable of eliciting
CC anti-major histocompatibility (MHC) antibodies. The composition
CC comprises an antigen associated with an altered MHC Class II determinant
CC (I) comprising alpha1, alpha2, beta1 and beta2 polypeptide domains
CC encoded by a mammalian MHC Class II locus covalently linked to form a
CC polypeptide (I) containing beta2, alpha2, alpha1 and beta1 domains in
CC sequence. The resulting Antigen-MHC complex is recognizable by the T cell
CC receptor. The compositions are used for immunisation against, or
CC treatment of, a wide range of autoimmune diseases, e.g. acquired immune
CC deficiency syndrome (AIDS), lupus erythematosus, multiple sclerosis,
CC thyroiditis, toxic shock, tumour and snakebite, depending on the nature
CC of antigen. (I) is also used to analyse functional interactions between
CC the various domains and for targeting lymphocyte receptors. Antibodies
CC against (I) are produced by usual methods of immunisation or cell fusion,
CC and may be humanised by standard methods. These antibodies are useful for
CC diagnosis (detection or purification of MHC gene products), therapy
CC (neutralising MHC on cell surfaces) and in the study of MHC and cellular
CC processes. AA333240 to AA33242 and AAV52840 to AAV52970 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINPEKL 8
1 SIINPEKL 8

RESULT 27
AAG77871
ID AAG77871 standard; Protein; 8 AA.

AC AAG77871;

DT 08-MAY-2002 (first entry)

DE Ovalbumin 257-264 peptide.

XX Ovalbumin 257-264 peptide; major histocompatibility complex;
KW MHC-peptide complex; MHC; human; MHC class I alpha chain;
KW beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;
KW vaccine; immune response modulation; hyperproliferative disorder;
KW neoplasm; hypergammaglobulinaemia; viral infection; hepatitis;
KW meningitis; bacterial infection; tuberculosis; gingivitis;
KW parasitic infection; autoimmune disease; Hashimoto's disease;
KW Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;
KW graft-versus-host disease; GVHD; antigenic peptide.

OS Aves.

PN WO200178768-A2.

PD 25-OCT-2001.

PF 12-APR-2001; 2001WO-US11912.
PR 12-APR-2000; 2000US-196472P.

XX (UVRP) UNIV ROCHESTER.

PI Zauderer M, Smith ES;

DR WPI; 2001-602927/68.

PT Novel compound comprising major histocompatibility complex-peptide
PT complexes, used to modulate immune responses -
XX Example 19; Page 101; 166pp; English.

XX The invention comprises a compound which contains one or more major
CC histocompatibility complex (MHC)-peptide complexes, and an antibody
CC specific for a cell surface marker. The complexes comprise an MHC class
CC I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide
CC bound in the MHC groove. Alternatively, the complexes may comprise an MHC
CC class II alpha chain, an MHC class II beta chain, and an antigenic
CC peptide bound in the MHC groove. The complexes are linked to the carboxyl
CC termini of the antibody. The compounds of the invention can be used as a
CC vaccine to modulate an immune response. The compounds of the invention
CC are useful for treating: hyperproliferative disorders (e.g. neoplasms and
CC hypergammaglobulinaemia); viral infections (e.g. hepatitis and
CC meningitis); bacterial infections (e.g. tuberculosis and gingivitis);
CC parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,
CC Graves' disease and rheumatoid arthritis); allergic reactions/conditions
CC (e.g. asthma). The compounds of the invention may also be used in the
CC treatment of organ rejection or graft-versus-host disease (GVHD). The
CC present amino acid sequence represents the ovalbumin 257-264 peptide,
CC which was used as an antigenic peptide in an example of the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY      1 SIINFEXL 8
XX      |||||
DB      1 SIINFEXL 8

RESULT 28
ID      AAM52562 standard; Peptide; 8 AA.
XX
XX      AAM52562;
AC
XX      AAM52562;
DT      04-FEB-2002 (first entry)
XX
DE      Cytotoxic T-cell epitope for ovalbumin.
XX
XX      Cell death; toxic gene; tumour suppressor; ovalbumin;
XX      Cytotoxic T-cell epitope.
XX
XX      Unidentified.
XX      OS
XX      MO200172995-A2.
XX
XX      PD
XX      04-OCT-2001.
XX
XX      PF
XX      28-MAR-2001; 2001WO-US09953.
XX
XX      PR
XX      28-MAR-2001; 2000US-0192586.
XX      10-MAY-2001; 2000US-020343.
XX      23-JAN-2001; 2001US-0263226.
XX      27-FEB-2001; 2001US-0271426.
XX
XX      PA
XX      (UNRP ) UNIV ROCHESTER.
XX      PI
XX      Zauderer M, Smith ES;
XX      WPI; 2001-570897/64.
XX
DR      WPI; 2001-570897/64.
XX
PT      Selecting target polynucleotides, particularly toxic genes, involves
PT      introducing a library of insert polynucleotides into a host cell
PT      population, where the target polynucleotide promotes cell death -
XX
XX      Example 1; Page 136; 359pp; English.
XX
PS      The present invention relates to a method for selecting a target
XX      polynucleotide. The method comprises introducing into a host cell
XX      a population a library of insert polynucleotides, where expression of the
XX      target polynucleotide directly or indirectly promotes host cell death.
XX      The cells are cultured and the insert polynucleotides are collected from
XX      the cells which die. The method is useful for selecting target
XX      polynucleotides, particularly polynucleotides which alter cell phenotypes
XX      of induce or inhibit cell death. The method can be used to isolate toxic
XX      genes such as tumour suppressors. The present sequence was used to
XX      illustrate the method of the the present invention.
XX
SQ      Sequence      8 AA;

Query Match      100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

OY      1 SIINFEXL 8
XX      |||||
DB      1 SIINFEXL 8

RESULT 29
ID      AAE13119 standard; peptide; 8 AA.
XX
XX      AAE13119;
AC
XX
XX      28-JAN-2002 (first entry)
DT

```

```

XX      Ovalbumin (OVA)-derived immunodominant octapeptide.
DE
XX      Immunostimulatory fusion protein; IFP; antigen component; therapy;
XX      immunostimulatory component; T-cell mediated immune response; DC;
XX      dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
XX      Ovalbumin; OVA-derived immunodominant octapeptide.
XX
XX      Unidentified.
XX      OS
XX      WO200174855-A2.
XX
XX      PD
XX      11-OCT-2001.
XX
XX      PF
XX      30-MAR-2001; 2001WO-US10515.
XX
XX      PR
XX      30-MAR-2001; 2000US-193504P.
XX
XX      PA
XX      (DENR-) DENDREON CORP.
XX
XX      PI
XX      Laus R, Vidovic D, Graddis T;
XX      WPI; 2001-662965/76.
XX
DR      WPI; 2001-662965/76.
XX
XX      PT
XX      An immunostimulatory fusion protein comprising the intracellular domain
XX      of HER-2 and an antigen elicits an immune response to the antigen and
XX      is useful for the treatment of associated cancer associated -
XX
XX      Example 1; Page 25; 59pp; English.
XX
PS      The invention relates to immunostimulatory fusion proteins (IFP) and
XX      nucleic acid molecules encoding such proteins. The IFPs comprise a
XX      polypeptide antigen component and an immunostimulatory component derived
XX      from the intracellular domain of HER-2 protein which is effective to
XX      elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
XX      immune response to the antigen. IFP or superactivated dendritic cells
XX      are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
XX      associated with a particularly antigen. The present sequence is a
XX      ovalbumin (OVA)-derived immunodominant octapeptide. This peptide
XX      is used in the fusion constructs of the invention.
XX
SQ      Sequence      8 AA;

Query Match      100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

OY      1 SIINFEXL 8
XX      |||||
DB      1 SIINFEXL 8

RESULT 30
ID      AAE12145 standard; peptide; 8 AA.
XX
XX      AAE12145;
AC
XX      AAE12145;
DT      15-JAN-2002 (first entry)
XX
DE      Murine ovalbumin (OVA) peptide.
XX
XX      Microbial delivery vehicle; prophylactic; immunisation; gene therapy;
XX      tumour; carcinoma; neurodegeneration; muscular atrophy; cytostatic;
XX      neuroprotective; antibacterial; insecticide; fungicide; antiviral;
XX      antiprotocozal; cytostatic; anti-inflammatory; murine; ovalbumin; OVA;
XX      listeriolysin O; LLO; MHC; major histocompatibility complex.
XX
XX      OS
XX      Mus sp.
XX      US6287556-B1.
XX
XX      PD
XX      11-SEP-2001.

```

```

XX 21-DEC-1999; 99US-0469197.
XX
XX 13-AUG-1998; 98US-0133914.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Portnoy DA, Higgins DE;
XX
XX WPI; 2001-647179/74.
XX
XX Vaccine for preventing (e.g. as immunizations) or treating (e.g. as
XX gene therapy) tumor, carcinoma, neurodegeneration or muscular atrophy,
XX comprises a non-virulent bacterium
XX
XX Example; Column 12; 14pp; English.
XX
CC The invention relates to microbial-based intracellular delivery of
CC agents to eukaryotic cells. The agents include microbial delivery
CC vehicles such as nonvirulent bacteria comprising a first gene
CC encoding a nonsecreted foreign cyclolysin operably linked to a
CC heterologous promoter and a second gene encoding a different
CC foreign agent. The foreign agent may be a nucleic acid or protein,
CC and is frequently bioactive in and therapeutic to the target
CC eukaryote. The vaccine comprising nonvirulent bacterium is useful
CC for prophylactics (e.g. as immunisations) and treatments (e.g. as
CC gene therapy) of e.g. tumour, carcinoma, neurodegeneration or
CC muscular atrophy. The present sequence is murine ovalbumin (OVA)
CC peptide. This sequence is used to examine the ability of
CC Escherichia coli expressing listeriolysin O (LLO) and an antigenic
CC protein to deliver the antigen to the cytosol of macrophages for
CC processing and presentation on MHC (major histocompatibility
XX complex) class I molecules.
XX
SQ Sequence 8 AA;

```

Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 SIINFEKL 8
   |||||
DB 1 SIINFEKL 8

```

RESULT 31
AAE09514
ID AAE09514 standard; peptide; 8 AA.
XX
AC AAE09514;
XX
XX 19-NOV-2001 (first entry)
XX
DE Human ovalbumin peptide.
XX
XX Mucin; cytosolic; immunostimulant; cell mediated immune response;
XX carcinoma; adenocarcinoma; breast cancer; dendritic cell; vaccine;
XX gene therapy; human; ovalbumin.
XX
XX Homo sapiens.
XX
XX OS
XX WO200157068-A1.
XX
XX PD 09-AUG-2001.
XX
XX 01-FEB-2001; 2001WO-AU00090.
XX
XX 01-FEB-2000; 2000AU-0005369.
XX 14-JUN-2000; 2000US-0593870.
XX
XX (AUST-) AUSTIN RES INST.
XX
XX Mckenzie IFC, Pietersz GA, Apostolopoulos V;
PI

```

XX DR WPI; 2001-541537/60.
XX
XX Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a
XX an epitope of the non-VNTR, non-leader region of a mucin -
XX
XX Disclosure; Page 14; 84pp; English.
XX
XX The patent discloses peptide or polypeptides capable of eliciting
XX an immune response, comprising an amino acid sequence corresponding
XX to an epitope of the non-central portion of varying numbers of an
XX amino acid motif (VNTR), non-leader region of a mucin. The peptides
XX of the invention, fusion proteins comprising the peptide and conjugate
XX compounds with carbohydrate polymers are used to induce a cell mediated
XX immune response against mucin in the prevention or treatment of
XX carcinoma, preferably adenocarcinoma, most preferably breast cancer.
XX They are also used to pulse dendritic cell for in vivo transfer and
XX use as a vaccine. They are also used in gene therapy. The present
XX sequence is ovalbumin peptide from human. This sequence is used for
XX the prediction of T-cell epitopes.
XX
SQ Sequence 8 AA;

```

Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 SIINFEKL 8
   |||||
DB 1 SIINFEKL 8

```

RESULT 32
AAG63855
ID AAG63855 standard; peptide; 8 AA.
XX
XX AAG63855;
XX
XX 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of an OVA-derived minimal CTL peptide.
XX
XX Opti; lipoprotein; adjuvant; type1 immune response; gp63;
XX leishmania major; leishmaniasis; TBC; leprosy; mycotin infection;
XX allergic asthma; autoimmune disease.
XX
XX Synthetic.
XX
XX OS
XX WO200160404-A2.
XX
XX PN
XX 23-AUG-2001.
XX
XX 13-FEB-2001; 2001WO-EP01673.
XX
XX 18-FEB-2000; 2000EP-0200589.
XX
XX (VLA-) VLAAWS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
XX PA
XX Revets H, Cornelis P, De Baetselier P;
XX
XX PI
XX WPI; 2001-522552/57.
XX
XX DR
XX Use of major Opti lipoprotein of Pseudomonas aeruginosa or its
XX functional fragments as adjuvant to obtain a Th1 type immune response
XX against heterologous antigen, for treating leishmaniasis, leprosy,
XX allergic asthma
XX
XX PS
XX Disclosure; Page 12; 54pp; English.
XX
XX The present sequence represents an OVA-derived minimal CTL peptide,
XX which was used in the course of the invention. The specification
XX describes the use of the major Opti lipoprotein of Pseudomonas
XX aeruginosa or its functional fragments as an adjuvant to obtain

CC a Th1 type immune response against a heterologous antigen. They are
 CC especially used as an adjuvant to obtain a Th1 type immune response
 CC against a heterologous antigen such as antigen gp63 of *Leishmania* major,
 CC for treating a disease such as leishmaniasis, TB (undefined), leprosy,
 CC mycobacter infection, allergic asthma or an autoimmune disease, in which
 CC the natural Th1 response is insufficient and/or in which the immune
 CC response is polarizes towards Th2 response.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 STINPEKL 8
 |||||
 1 STINPEKL 8

RESULT 33

AA66422 standard; peptide; 8 AA.

AC AA66422;

DT 23-OCT-2001 (first entry)

DE Chicken ovalbumin peptide, OVA257-264, used as a peptide antigen.

XX Immunomodulator; vaccine; immune response; immunogenic; chicken;

KW ovalbumin.

XX Gallus domesticus.

PN WO200154720-A1.

PD 02-AUG-2001.

PS 05-JAN-2001; 2001WO-EP00087.

PR 28-JAN-2000; 2000AT-0000129.

PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PI Lingnan K, Matner F, Schmidt W, Birnstiel M, Buschle M;

DR WPI; 2001-536419/59.

PT Pharmaceutical composition useful for inducing immune response

PT comprises antigen, immunogenic oligodeoxynucleotide containing

PT cytosine-guanine dinucleotide motifs and polycationic polymer -

XX Example 1; Page 22; 39pp; English.

XX The present invention relates to a pharmaceutical composition which

XX comprises an antigen, an immunogenic oligodeoxynucleotide containing

XX cytosine-guanine dinucleotide (CpG) motifs (CpG-ODN) and a polycationic

XX polymer. The composition is useful for making a vaccine to induce potent

XX immune responses, or to decrease or ablate undesired immune responses.

XX The present sequence, OVA257-264, is a peptide from chicken ovalbumin.

XX This sequence was used as a peptide antigen in the method of the present

XX invention.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 STINPEKL 8
 |||||
 1 STINPEKL 8

RESULT 34

AA05398 standard; peptide; 8 AA.

AC AA05398;

DT 24-OCT-2001 (first entry)

DE Peptide released from ovalbumin (Ova) after cleavage of peptide P1.

XX Heat shock protein; hsp; CD8+ cytotoxic T lymphocyte; ovalbumin;

XX CTL; CD4+ T cell; AIDS; acquired immunodeficiency syndrome; murine;

XX human immunodeficiency virus; HIV; pathogen; cancer.

OS Mus sp.

PN WO200151081-A1.

PD 19-JUL-2001.

PS 01-DEC-2000; 2000WO-US32831.

PR 14-JAN-2000; 2000US-0176143.

PA (WHEB) WHITEHEAD INST BIOMEDICAL RES.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Huang Q, Richmond JFL, Cho BK, Palliser D, Chen J, Eisen HN;

DR WPI; 2001-451815/48.

PT Inducing a CD8+ cytotoxic T lymphocyte immune response in an individual

PT for treating diseases such as HIV involves administering a fusion

PT molecule comprising a heat shock protein -

XX Example 1; Fig 1C; 58pp; English.

XX The present sequence represents a naturally occurring murine peptide

XX which is released from ovalbumin (Ova) upon cleavage of peptide P1

XX (AA05397). The present sequence is described in an invention

XX relating to a novel method of inducing a CD8+ cytotoxic T lymphocyte

XX (CTL) response to a molecule in an individual by administering a

XX fusion molecule joined to a hsp, or an adenosine triphosphate (ATP)

XX binding domain of a hsp. The method is particularly useful in inducing

XX a CD8+ CTL response in an individual deficient in CD4+ T cells e.g. for

XX treating an AIDS acquired immunodeficiency syndrome patient carrying

XX the human immunodeficiency virus (HIV). The method is also useful for

XX treating diseases that are caused by or associated with intracellular

XX pathogens, and for treating cancer.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 STINPEKL 8
 |||||
 1 STINPEKL 8

RESULT 35

AA06033 standard; peptide; 8 AA.

AC AA06033;

DT 25-SEP-2001 (first entry)

DE Chicken ovalbumin CTL epitope.

KM Chicken; vaccine; cytostatic; immunostimulant; antibacterial; antifungal;
 KM protozoacide; antiviral; recombinant flavivirus; immune response; TAA;
 KM tumour associated antigen; ovalbumin; epitope.
 XX
 OS Gallus sp.
 XX
 PN WO200153467-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 19-JAN-2001; 2001WO-US01866.
 XX
 PR 21-JAN-2000; 2000US-0177449.
 PR 01-SEP-2000; 2000US-0653754.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Andino-pavlovsky R, Mcallister-moreno A;
 PI WPI; 2001-457605/49.
 DR
 XX
 PT Recombinant flavivirus useful for reducing the number of tumor cells in
 PT a host comprises an exogenous nucleic acid encoding a polypeptide
 PT proteolytically processed after expression in the vector resulting in
 PT its release -
 PS
 XX Example 1; Page 19; 48pp; English.
 XX
 CC The invention relates to a replication-competent recombinant flavivirus
 CC (yellow fever virus-YFV) comprising an exogenous nucleic acid encoding an
 CC exogenous polypeptide. The replication-competent recombinant flavivirus
 CC may be used to illicit an immune response to an antigen, e.g., tumour
 CC associated antigen (TAA) or a microbial pathogen antigen, in a mammalian
 CC host. Particularly, recombinant flavivirus is useful for reducing the
 CC number of tumour cells in a host. In addition, recombinant flavivirus
 CC may be useful as a vaccine to provide immune protection against pathogens
 CC such as bacteria, viruses, fungi and parasites. Unlike other vectors
 CC which will produce only one cycle of antigen expression and/or which
 CC will stop expression without the intervention of the host immune system,
 CC recombinant flaviviruses continue to propagate until the immune system
 CC is sufficiently activated to halt the infection. This produces a stronger
 CC immune response against the exogenous antigenic peptide produced from
 CC the flavivirus as compared to the immune response that would be elicited
 CC using conventional expression vectors (e.g., a viral replicon). The
 CC present peptide sequence is chicken ovalbumin CTL epitope which is the
 CC exogenous polypeptide.
 CC
 SQ Sequence 8 AA;
 XX
 OY
 DB 1 SIINFEKL 8
 1 SIINFEKL 8
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 36
 AAB84316
 ID AAB84316 standard; peptide; 8 AA.
 XX
 AC AAB84316;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Peptide used to produce lemA peptide variants.
 XX
 XX lemA; CD8+ epitope; T cell response.
 XX
 OS Synthetic.
 XX

PN WO200140275-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 06-DEC-2000; 2000WO-US33027.
 XX
 PR 06-DEC-1999; 99US-0169227.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Kurlander RJ, Chao E, Fields J;
 PI WPI; 2001-389952/41.
 DR
 XX
 PT New isolated variant of lemA, lemA, comprising a hydrophobic element
 PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -
 PS
 XX Disclosure; Page 6; 65pp; English.
 XX
 CC The specification describes a peptide variant of lemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC can be inserted near the lemA hydrophobic element, to create peptides
 CC of the invention.
 CC
 SQ Sequence 8 AA;
 XX
 OY
 DB 1 SIINFEKL 8
 1 SIINFEKL 8
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 37
 AAB93354
 ID AAB93354 standard; Peptide; 8 AA.
 XX
 AC AAB93354;
 XX
 DT 24-AUG-2001 (first entry)
 XX
 DE Ovalbumin cytotoxic T lymphocyte epitope SEQ ID NO:11.
 XX
 DE Immunostimulatory sequence; ISG; immunomodulatory; immune response;
 KM antigen; anti-allergic; modulation; Th1 lymphocyte stimulation; allergy;
 KM Th1-associated cytokine; Th2 lymphocyte suppression; cytokine.
 XX
 OS Synthetic.
 XX
 PN WO200135991-A2.
 XX
 PD 25-MAY-2001.
 XX
 PF 15-NOV-2000; 2000WO-US31385.
 XX
 PR 15-NOV-1999; 99US-0165467.
 PR 14-NOV-2000; 2000US-0713136.
 XX
 PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
 XX
 PI Tuck S, Van Nest G;
 PI WPI; 2001-329209/34.
 DR
 XX
 PT Populations of conjugate molecules comprising polynucleotide
 PT immunostimulatory sequences polynucleotides and antigens, useful for

PT controlling immune responses -
 XX
 PS Example 6; Page 69; 97pp; English.
 XX
 CC The present invention describes immunomodulatory populations ((i) and
 CC ((ii)) of conjugate molecules (CMs) comprising immunostimulatory sequences
 CC ((ISS)) of polynucleotides and antigens. The extent of conjugation affects
 CC the immunological properties (e.g. the extent of antigen-specific
 CC antibody formation, including Th1-associated antibody formation) so the
 CC conjugates are used for altering the type and extent of immune response.
 CC ((i) and ((ii)) have immunomodulatory, immunosuppressive and anti-allergic
 CC activities, and can be used in the modulation of immune responses via
 CC the stimulation of Th1 lymphocytes and Th1-associated cytokines, and
 CC suppression of Th2 lymphocytes and cytokines. The populations ((i) and
 CC ((ii)) of conjugate molecules may be used for modulating immune responses
 CC in individuals e.g. for the treatment of an allergic condition. ((i) and
 CC ((ii)) may be used to modulate immune responses and therefore prevent
 CC potentially harmful reactions to antigens. The present sequence
 CC represents an ovalbumin (OVA) cytotoxic T lymphocyte (CTL) epitope
 CC which is used in the exemplification of the present invention.
 XX
 SQ Sequence 8 AA:
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 RESULT 38
 AAB82176
 ID AAB82176 standard; peptide; 8 AA.
 XX
 AC AAB82176;
 XX
 DT 20-JUL-2001 (first entry)
 XX
 DE Immunodominant CTL epitope of ovalbumin.
 XX
 KW Vaccine; Antiviral; Antibacterial; Antiparasitic; liposome;
 KW archaeobacterium; cytotoxic T lymphocyte response;
 KW immunodominant epitope; ovalbumin; archaeosome.
 XX
 OS Unidentified.
 XX
 PN WO200126683-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 12-OCT-2000; 2000WO-CA01197.
 XX
 PR 12-OCT-1999; 99US-0158944.
 PR 08-JUN-2000; 2000US-0209988.
 XX
 PA (CANA) NAT RES COUNCIL CANADA.
 XX
 PI Sprott GD, Krishnan L, Conlan JW, Omri A, Patel GB;
 DR WPI; 2001-261839/29.
 XX
 PT New vaccine comprising a liposome useful for conferring protective
 PT immunity against an intracellular pathogen -
 XX
 PS Disclosure; Page 34; 96pp; English.
 XX
 CC The present invention relates to a vaccine composition comprising a
 CC liposome prepared from the total polar lipid extract of an
 CC archaeobacterium and an acellular antigen, preferably an isolated
 CC outer membrane from a pathogen. The vaccine of the invention provides an
 CC enhanced cytotoxic T lymphocyte response. The vaccine of the invention

CC is useful for conferring protective immunity against an intracellular
 CC pathogen. The present peptide: immunodominant CTL epitope of ovalbumin,
 CC was used to illustrate the present invention. This peptide was used to
 CC test for the ability of archaeosomes to induce CTL responses to
 CC ovalbumin.
 XX
 SQ Sequence 8 AA:
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 RESULT 39
 AAB81122
 ID AAB81122 standard; peptide; 8 AA.
 XX
 AC AAB81122;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Chicken ovalbumin (OVA) peptide.
 XX
 KW Ovalbumin; chicken; OVA; immune response; Langerhans cell migration;
 KW tumour; EGF-OVA.
 XX
 OS Gallus gallus.
 XX
 PN US6210672-B1.
 XX
 PD 03-APR-2001.
 XX
 PF 20-OCT-1998; 98US-0176044.
 XX
 PR 20-OCT-1998; 98US-0176044.
 XX
 PA (TORR-) TORREY PINES INST MOLECULAR STUDIES.
 XX
 PI Cowing C;
 DR WPI; 2001-280845/29.
 XX
 PT Enhancing an immune response to an antigen in a mammal comprises
 PT topically administering the antigen, a penetration enhancer and an
 PT agent for enhancing Langerhans cell migration -
 XX
 PS Disclosure; Column 4; 15pp; English.
 XX
 CC This invention relates to a method for enhancing an immune response to an
 CC antigen in a mammal. The method comprises administering a composition
 CC comprising the antigen, a penetration enhancer selected from lipophilic
 CC solvents, low-frequency ultrasound, electroporation, iontophoresis and
 CC intraepidermal delivery and an agent for enhancing Langerhans cell
 CC migration to an epidermal or mucosal site. The method can be used to
 CC enhance the immune response to tumours, viruses, bacteria and parasites.
 CC The present sequence represents a fragment of the chicken ovalbumin (OVA)
 CC protein. The peptide functions as a EGF-OVA tumour associated peptide
 CC antigen for CD8+ cytotoxic T lymphocytes. The peptide can be used in the
 CC method of the invention to enhance an immune response to the EGF-OVA
 CC tumour.
 XX
 SQ Sequence 8 AA:
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8

```

DB                               1 SIINFEKL  8

RESULT 40
AAB82065
ID  AAB82065 standard; peptide; 8 AA.
AC  AAB82065;
XX
XX
XX  22-JUN-2001 (first entry)
XX
XX  Ovalbumin-derived peptide, used as a control peptide.
DE
XX  Antigen; immunostimulant; vaccine; pharmaceutical composition; antiviral;
KM  viral infection; ovalbumin.
XX
XX  unidentified.
OS
XX
XX  WO200124822-A2.
PN  12-APR-2001.
PD
XX
XX  02-OCT-2000; 2000MO-EP09657.
PF
XX
XX  01-OCT-1999; 99AT-0001680.
PR
XX
XX  (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
PA
XX
XX  Fleitmann J, Mattner F, Buschle M, Welling J;
PI  WPI, 2001-290577/30.
XX
XX
XX  New pharmaceutical composition comprising an antigen, an
PT  immunostimulating substance and a polycationic polymer, useful in
PT  manufacturing vaccines -
XX
XX
XX  Example 1; Page 11; 20pp; English.
XX
XX
XX  The present invention relates to a pharmaceutical composition comprising
CC  (a) an antigen; (b) an immunostimulating substance consisting of
CC  neuroactive compounds, hormones, compounds having growth hormone activity
CC  or their mixtures; and (c) a polycationic polymer. The composition is
CC  useful in manufacturing vaccines. To illustrate the present invention, a
CC  murine tyrosinase related protein-2 peptide (TRP-2 peptide; see
CC  AAB82064), was used. Mice were injected subcutaneously with either the
CC  TRP-2 peptide, TRP-2 peptide + human growth hormone (hGH), TRP-2 peptide
CC  + poly-L-arginine 60 (pR60) or TRP-2 peptide + pR60 + hGH. Animals were
CC  sacrificed 10 days post injection, and mesenteric and inguinal lymph
CC  nodes were harvested. Lymphocytes were prepared from lymph nodes and were
CC  re-stimulated with TRP-2 peptide or with an ovalbumin-derived peptide
CC  (the present peptide), with the same major histocompatibility complex
CC  (MHC) restriction serving as negative control. Spots representing single
CC  T cells specific for the peptide used for re-stimulation were counted. No
CC  spots were detected when the ovalbumin derived peptide was used, while
CC  TRP-2 peptide + pR60 + hGH showed the highest number of spots or single T
CC  cells. The present peptide was also used as a control peptide for
CC  experiments with substance P (see AAB82070).
XX
XX
SQ  Sequence      8 AA;

Query Match      100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 SIINFEKL  8
        |||||
DB      1 SIINFEKL  8

RESULT 41
AAB92374
ID  AAB92374 standard; peptide; 8 AA.
XX

```

AC	AA892374;
XX	
DT	22-JUN-2001 (first entry)
XX	
DE	Miscellaneous peptide SEQ ID NO:1550.
KX	Protection; endogenous therapeutic peptide; peptidase; conjugation;
KW	blood component; modification; succinimidy; maleimido group; amino;
KW	hydroxyl; thiol; hormone; growth factor; neurotransmitter.
XX	
OS	Homo sapiens.
OS	Synthetic.
PN	WO200069900-A2.
PD	
PE	23-NOV-2000.
PF	17-MAY-2000; 2000WO-US13576.
PR	17-MAY-1999; 99US-0134406.
PR	10-SEP-1999; 99US-0153406.
PR	15-OCT-1999; 99US-0159783.
XX	
PA	(CONU-) CONJUCHEM INC.
PI	Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudau K;
DR	WPI; 2001-112059/12.
PT	
PT	Modifying and attaching therapeutic peptides to albumin prevents
PT	peptide degradation, useful for increasing length of in vivo activity
PS	-
PS	Disclosure; Page 711; 733pp; English.
CC	The present invention describes a modified therapeutic peptide (I)
CC	comprising a therapeutically active amino acid region (III) and a
CC	reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
CC	a less therapeutically active amino acid region (IV), which covalently
CC	bonds with amino/hydroxyl/thiol groups on blood components to form a
CC	peptide stabilised therapeutic peptide composed of 3-50 amino acids.
CC	(I) are useful for modifying therapeutic peptides e.g. hormones, growth
CC	factors and neurotransmitters, to protect them from peptidase activity
CC	in vivo for the treatment of various disorders. Endogenous therapeutic
CC	peptides are not suitable as drug candidates as they require frequent
CC	administration due to rapid degradation by peptidases in the body.
CC	Modifying and attaching therapeutic peptides to albumin prevents or
CC	reduces the action of peptidases to increase length of activity (half
CC	life) and specifically as bonding to large molecules decreases
CC	intracellular uptake and interference with physiological processes.
CC	AA890829 to AA892441 represent peptides which can be used in the
CC	exemplification of the present invention.
XX	
SQ	Sequence 8 AA:
OY	
DB	
QY	Query Match 100.0%; Score 38; DB 22; Length 8;
	Best Local Similarity 100.0%; Pred. No. 9.3e+05;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	1 SIINFEKL 8
	1 SIINFEKL 8
XX	
AC	AA848950 standard; Protein; 8 AA.
DT	27-MAR-2001 (first entry)
XX	
DE	Ovalbumin MHC class I epitope, SEQ ID NO:6.

```

XX Transposable element; MHC epitope; major histocompatibility complex;
KM intracellular bacterial pathogen; loxp site; Cre recombinase;
KM insertion end; in-frame fusion; detection; antigen;
KM diseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
KM ovalbumin MHC class I epitope.
XX Unidentified.
XX MO200071158-A1.
XX 30-NOV-2000.
XX 26-MAY-2000; 2000WO-US14687.
XX 26-MAY-1999; 99US-0136210.
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX Heffron FL, Parker DC, Ellefson DB;
XX WPI; 2001-031967/04.
XX Transposable element for detecting an antigenic epitope of a pathogen,
PT comprising 5' and 3' recombining sites, nucleic acid sequences encoding
PT a selectable marker and major histocompatibility complex (MHC) epitope,
PT and an insertion end -
XX Claim 5; Page 41; 63pp; English.
XX The invention relates to a novel transposable element comprising DNA
XX encoding a selectable marker (e.g., antibiotic resistance) located
XX between a 5' recombining site and a 3' recombining site (e.g., loxp
XX sites); DNA encoding an MHC (major histocompatibility complex) epitope
XX either 5' of the 5' recombining site or 3' of the 3' recombining site;
XX and insertion ends comprising an inverted repeat sequence at the 5' and
XX 3' ends of the transposable element sufficient for integration of the
XX transposable element. The transposable elements of the invention are able
XX to introduce in-frame insertions throughout the chromosome of an
XX intracellular bacterial pathogen. This system "tags" the bacterial gene
XX and resulting protein, allowing the identification of proteins
XX secreted across the membranes of the eukaryotic cell infected by the
XX bacterium. In one embodiment, the transposable elements contain an
XX antibiotic resistance cassette, two minimal loxp recombination sites, an
XX MHC class I or class II epitope, and flanking insertion ends. A
XX transposase, such as the Cre recombinase protein, is expressed in trans
XX from a plasmid, or can be included in the transposable element. The Cre
XX recombinase loops out the intervening sequences containing the antibiotic
XX resistance cassette. When the transposable element inserts within a gene,
XX the resolved insertion places the MHC class I or class II epitope in
XX frame with the gene. The transposable elements of the invention are
XX useful for detecting an antigenic epitope of an intracellular bacterial
XX pathogen, such as Salmonella sp., Mycobacterium tuberculosis and Listeria
XX monocytogenes. Certain embodiments of the technology, termed
XX "diseminated insertions of class-I epitopes" (DICE-I; DICE-II for
XX class II epitopes) allow the rapid and accurate identification of
XX proteins involved in bacterial pathogenesis so that such proteins can
XX be used as vaccine and drug targets. Carrier vaccines may be generated
XX by infecting bacteria with a transposable element of the invention
XX which additionally comprises an antigen associated with a disease,
XX preferably cancer or a viral or bacterial disease, operably linked to the
XX MHC epitope DNA of the transposable element. The present sequence
XX represents an ovalbumin MHC class I epitope specifically claimed
XX for use in the invention.
XX
SQ Sequence 8 AA;

```

```

Query Match          100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SIINFEKL 8
|||||
XX

```

```

DB 1 SIINFEKL 8
RESULT 43
AAE28959
ID AAE28959 standard; peptide; 8 AA.
XX
XX AAE28959;
XX
XX 27-JAN-2003 (first entry)
XX
XX Chicken ovalbumin peptide.
XX
XX Modified vaccinia Ankara virus; MVA; HIV; human immunodeficiency virus;
KM CD8+ T cell; immune response; acquired immune deficiency syndrome; AIDS;
KM viral infection; vaccine; immunostimulant; virucide; chicken.
XX
XX Gallus sp.
XX
XX WO200272754-A2.
XX
XX 19-SEP-2002.
XX
XX 01-MAR-2002; 2002WO-US06713.
XX
XX 08-MAR-2001; 2001US-274434P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Moss B, Wyatt L, Earl P;
XX
XX WPI; 2002-723330/78.
XX
XX New recombinant modified vaccinia Ankara (MVA) virus expressing HIV
XX env, gag and pol genes, useful for boosting or inducing CD8 T cell
XX immune responses in primates, e.g. humans, particularly for preventing
XX AIDS or other viral infections -
XX
XX Example 1; Page 111; 112pp; English.
XX
XX The present invention relates to a composition comprising recombinant
XX modified vaccinia Ankara (MVA) virus expressing an HIV (human immuno-
XX deficiency virus) env, gag and pol gene or its modified gene for the
XX production of an HIV Env, Gag and Pol antigen by expression from the
XX recombinant MVA virus. The HIV env gene is modified to encode an HIV
XX Env protein composed of gp120 and the membrane-spanning and ectodomain
XX of gp41, but lacking part or all of the cytoplasmic domain of gp41. The
XX composition or recombinant MVA virus is useful for boosting or inducing
XX CD8+ T cell immune response in primates, particularly in humans. The
XX composition may be used for preventing AIDS (acquired immune deficiency
XX syndrome) or other viral infections. Sequences of the invention are also
XX used as vaccines. The present sequence is chicken ovalbumin peptide. This
XX sequence is used in the exemplification of the invention.
XX
SQ Sequence 8 AA;

```

```

Query Match          100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SIINFEKL 8
|||||
DB 1 SIINFEKL 8

```

```

RESULT 44
AAE26368
ID AAE26368 standard; peptide; 8 AA.
XX
XX AAE26368;
XX
XX 13-DEC-2002 (first entry)
XX

```

DE Ovalbumin CTL epitope.
 XX
 XX Human; immune response; T-helper cell epitope; chitosan; CTL response;
 KM vaccine; prostate cancer; breast cancer; cytostatic; immunostimulant.
 XX
 OS Unidentified.
 XX
 PN WO200234287-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 26-OCT-2001; 2001WO-DK00705.
 XX
 PR 27-OCT-2000; 2000DK-0001606.
 PR 03-NOV-2000; 2000US-245166P.
 PR 18-JUN-2001; 2001DK-0000936.
 PA (PHAR-) PHARMEXA AS.
 XX
 PI Beier AM, Gautam A, Mouritsen S;
 XX
 DR WPI; 2002-463339/49.
 XX
 PT Inducing or enhancing an immune response against an antigen,
 PT particularly cytotoxic T-lymphocyte responses, for treating or
 PT ameliorating prostate or breast cancer, comprises administering the
 PT antigen formulated with chitosan -
 XX
 PS Example 3; Page 63; 97pp; English.
 XX
 CC The invention relates to a method for inducing or enhancing an immune
 CC response against a polypeptide antigen in an animal, including human.
 CC The method comprises administering the polypeptide antigen or at least
 CC one variant which includes at least one first T-helper cell epitope that
 CC is foreign to the animal (foreign TH epitope) and is formulated with
 CC chitosan. The polypeptide antigen is weakly immunogenic or non-
 CC immunogenic. The invention is used as vaccine. The chitosan and
 CC polypeptide antigen or its variant are useful in the preparation of an
 CC immunogenic composition for inducing or enhancing an immune response,
 CC particularly CTL response, against the polypeptide or protein antigen.
 CC The method for inducing or enhancing an immune response is useful in
 CC treating or ameliorating cancer, e.g. prostate or breast cancer. The
 CC present sequence is ovalbumin CTL epitope used to illustrate the method
 CC of the invention.
 CC
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 XX
 AC ABB79933 standard; Peptide; 8 AA.
 XX
 AC ABB79933;
 XX
 DT 12-DEC-2002 (first entry)
 XX
 DE Ovalbumin T-cell epitope.
 KM Vaccine; genetic immunisation; gene therapy; antigen; epitope;
 KM T-cell; T-lymphocyte; ovalbumin.
 XX
 OS Unidentified.
 OS
 XX
 PN US2002115625-A1.
 XX

PD 22-AUG-2002.
 XX
 XX 08-MAR-2001; 2001US-0801540.
 PF
 XX 19-MAY-1999; 99US-0308511.
 PR
 XX (BONA/) BOT A.
 PA (BONA/) BONA C.
 XX
 PI Bot A, Bona C;
 XX
 DR WPI; 2002-712482/77.
 XX
 PR Immunizing an infant mammal against a target antigen or inducing a
 PR cytotoxic T cell response against a pathogen in the mammal, comprises
 PT administering nucleic acid encoding relevant epitopes of pathogen
 PT associated target antigens -
 XX
 PS Disclosure; Page 4; 45pp; English.
 XX
 CC The present sequence is that of an ovalbumin T-cell epitope (amino
 CC acids 257-264). This is a T-CTL epitope which, in the context of
 CC MHC Class II self antigens, may be recognised by a cytotoxic T-cell
 CC and thereby promote CTL-mediated lysis of cells comprising the target
 CC antigen. It is an example of T-cell epitopes which may be used
 CC according to the invention. The invention relates to immunising an
 CC infant mammal against a target antigen or inducing a cytotoxic
 CC T-cell response against a pathogen. The method involves inoculating
 CC the infant with a nucleic acid encoding one or more relevant epitopes
 CC of one or more target antigens associated with the pathogen in a
 CC carrier, so that the relevant epitope(s) is expressed in the infant
 CC mammal. B- or T-cell epitopes may be used, and the pathogen may be
 CC a virus, bacterium, protozoan, fungus, yeast, or parasite. The method
 CC may reduce the need for subsequent boost administrations and may
 CC prevent the side-effects associated with live attenuated vaccines.
 CC Administration of multiple epitopes directed to antigens
 CC associated with more than one pathogen may provide an infant with a
 CC broader spectrum of protection, and may be a means for inducing an
 CC immune response to a variety of childhood pathogens.
 CC
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 XX
 AC ABG93028 standard; Peptide; 8 AA.
 XX
 AC ABG93028;
 XX
 DT 20-NOV-2002 (first entry)
 XX
 DE Mouse class I MHC molecule Kb binding ovalbumin epitope.
 XX
 KW Regulator; transcription; cell death; phenotype; molecular scaffold;
 KW gene therapy; cancer; cardiovascular disease; arthritis; heart failure;
 KW ischaemia; obesity; neurodegenerative disease; Alzheimer's disease;
 KW bone pathology; dermatologic disease; psoriasis; infection; AIDS;
 KW acquired immunodeficiency syndrome; cosmetic; wound healing;
 KW antibiotic transport; drug toxicity; drug resistance; immunobiology;
 KW inflammation; allergic response; human immunodeficiency virus.
 XX
 OS Unidentified.
 OS
 XX
 PN WO200262822-A2.
 XX

PD 15-AUG-2002.
 XX
 PF 04-FEB-2002; 2002WO-US02814.
 XX
 PR 02-FEB-2001; 2001US-265586P.
 PR 05-FEB-2001; 2001US-265860P.
 PR 27-FEB-2001; 2001US-271423P.
 PR 23-JAN-2001; 2001US-263226P.
 PR 28-MAR-2000; 2000US-192586P.
 PR 22-SEP-1997; 97US-935377P.
 PA (UVRP) UNIV ROCHESTER.
 PI Zauderer M, Smith ES;
 XX
 DR WPI; 2002-643398/69.
 XX
 PT Identifying regulator polypeptides which influence target
 PT transcriptional regulatory regions, useful for treating cancer,
 PT comprises introducing host cells expressing the polypeptide into a
 PT library of polynucleotides
 XX
 PS Example 1; Page 82; 224pp; English.
 XX
 CC The invention discloses a method for identifying polynucleotides encoding
 CC a regulator polypeptide, whose expression induces activation of a target
 CC transcriptional regulatory region in a host cell. The method comprises
 CC providing a population of eukaryotic host cells capable of expressing the
 CC polypeptide, introducing into the host cell a library of polynucleotides
 CC encoding the polypeptides, permitting expression of the polypeptides and
 CC then recovering them from the host cells. The target transcriptional
 CC regulatory region is operably associated with a polynucleotide encoding a
 CC gene product, the expression of which results in host cell death or cause
 CC the host cells to exhibit a pre-determined modified phenotype and where
 CC the gene product is expressed upon activation of target transcriptional
 CC regulatory region. Each candidate regulator polypeptide comprises a
 CC candidate peptide and a molecular scaffold fused to the peptide so that
 CC the peptide is displayed on the surface of the candidate regulator
 CC polypeptide. The methods are useful in selecting and/or screening
 CC regulator molecules, such as polypeptides, which directly or indirectly
 CC induce or suppress the transcriptional activation of a target
 CC transcriptional regulatory region in a eukaryotic host cell. These
 CC regulator molecules may be used (e.g. in gene therapy) for preventing or
 CC treating cancers (e.g. breast or ovarian cancer), cardiovascular diseases
 CC (e.g. arhythmia, heart failure, ischaemia), obesity, neurodegenerative
 CC diseases (e.g. Alzheimer's disease), bone pathologies, dermatologic
 CC diseases (e.g. psoriasis), infections (e.g. viral, bacterial), acquired
 CC immunodeficiency syndrome (AIDS), in cosmetic applications and in wound
 CC healing. The method is also useful in screening regulator molecules that
 CC block antibiotic transport mechanisms, in drug toxicities and drug
 CC resistance applications and in improving the performance of existing or
 CC developmental drugs. It may also be used in immunobiology, inflammation,
 CC allergic response and in biotechnology applications. The sequences
 CC presented in ABG92946-ABG93029 are examples of regulator polypeptides.
 CC
 SQ Sequence 8 AA;
 XX
 SO
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STINFEKL 8
 |||||
 DB 1 STINFEKL 8
 |||||
 AC AAE25400;
 XX AAE25400 standard; peptide; 8 AA.
 XX
 DT 30-OCT-2002 (first entry)
 XX

XX
 DE Ovalbumin peptide used in the invention.
 XX
 KW Recombinant vector; coat protein; CP; viral replication; infection;
 KW Zucchini yellow mosaic potyvirus; ZYMV; cucurbit fruit; vaccination;
 KW pharmaceutical; diagnostic; ovalbumin.
 XX
 OS unidentified.
 PN WO200244323-A2.
 PD 06-JUN-2002.
 XX
 PF 28-NOV-2001; 2001WO-IL01098.
 XX
 PR 28-NOV-2000; 2000US-253136P.
 PR 27-SEP-2001; 2001US-0963761.
 XX
 PA (VIRO-) VIROGENE LTD.
 PI Gal-On A, Shioleth YM, Arazi T, Ilan Y;
 XX
 DR WPI; 2002-537446/57.
 DR N-PSDB; AAD41429.
 XX
 PT Novel recombinant vector useful for transiently expressing heterologous
 PT peptide in plant comprises potyvirus nucleic acid sequence and
 PT heterologous sequence inserted at amino terminus of potyvirus coat
 PT protein -
 XX
 PS Claim 17; Page 60; 61pp; English.
 XX
 CC The invention relates to a recombinant vector for expressing a
 CC heterologous peptide at the amino-terminus of a potyvirus coat protein
 CC (CP). The vector includes sufficient potyvirus nucleic acid sequence
 CC to permit viral replication and spread within a plant infected by the
 CC vector. The invention also relates to Zucchini yellow mosaic potyvirus
 CC (ZYMV) Agri strain CP and its corresponding nucleic acid sequence. The
 CC recombinant vector is useful for transiently expressing a portion of
 CC the heterologous peptide in a plant. It is also useful for infecting a
 CC cucurbit fruit, is useful as a source of material for vaccination,
 CC pharmaceutical or diagnostic application. The present sequence is a
 CC ovalbumin peptide used to fuse to the N-terminus of ZYMV Agri
 CC strain CP.
 XX
 SQ Sequence 8 AA;
 XX
 SO
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STINFEKL 8
 |||||
 DB 1 STINFEKL 8
 |||||
 AC ABG31661;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE Chicken ovalbumin MHC class I restricted epitope.
 XX
 KW Chicken; ovalbumin; MHC class I; major histocompatibility complex;
 KW polycationic compound; allergen; cytokine; chemokine; wound healing;
 KW cytotoxic drug; anti-oligogenic drug; immunostimulant; antiallergic;
 KW cytostatic; vulnary; immunogenic.
 XX
 OS Gallus gallus.
 XX

```

PN WO200253184-A2.
XX
XX 11-JUL-2002.
XX
XX 07-JAN-2002; 2002WO-EP00062.
XX
XX 05-JAN-2001; 2001WO-EP00087.
XX
XX 25-APR-2001; 2001AT-0000672.
XX
XX (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Lingnau K, Maltner F, Schmidt W, Birnstiel M, Buschle M;
XX WPI; 2002-627324/67.
XX
XX Use of a polycationic compound for the preparation of a medicament with
XX retarded in vivo release
XX
XX Example 1; Page 13; 29pp; English.
XX
XX The invention relates to preparation of a medicament with retarded in
XX vivo release comprising use of a polycationic compound. The compound is
XX used in the preparation of a medicament with retarded in vivo release,
XX and a vaccine containing an antigen. The medicament includes e.g. an
XX allergen, a cytokine, a chemokine, an immunostimulatory nucleic acid, a
XX cytotoxic or an anti-oligogenic drug or a compound needed for wound
XX healing. The medicament prevents or ameliorates side effects of drugs,
XX which are due to its too fast distribution of the drug throughout the
XX body by exhibiting a retarded release of the drug from the site of
XX administration. In the case of vaccine the compounds provide a depot,
XX which allows a long lasting continuous and effective presentation of the
XX antigen to the immune system to create a protective immunity. This
XX sequence represents a chicken ovalbumin MHC class I restricted epitope
XX used in the scope of the invention.
XX
XX Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 SIINFEKL 8
DB 1 SIINFEKL 8
XX
XX RESULT 49
XX ABG31967
XX ID ABG31967 standard; Peptide; 8 AA.
XX
XX AC ABG31967;
XX
XX DT 05-NOV-2002 (first entry)
XX
XX DE Chicken ovalbumin OVA257-264-peptide.
XX
XX KW Chicken; ovalbumin; polycationic; antiinflammatory; immunostimulant;
XX anti-allergic; cycostatic; vulnerary; medicament; inflammatory potential;
XX inflammation; vaccine; antigen; adjuvant; allergen; cytokine; chemokine;
XX immunostimulatory nucleic acid; cytotoxic drug; antioligogenic drug;
XX wound healing; OVA-peptide; epitope; major histocompatibility complex;
XX MHC.
XX
XX OS Gallus gallus.
XX
XX PN WO200253185-A2.
XX
XX PD 11-JUL-2002.
XX
XX PE 07-JAN-2002; 2002WO-EP00071.
XX
XX PF 05-JAN-2001; 2001WO-EP00087.
XX
XX PR

```

```

PR 25-APR-2001; 2001AT-0000670.
XX
XX (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Lingnau K, Eyed A, Schmidt W, Buschle M, Grill S;
XX WPI; 2002-627325/67.
XX
XX Use of a polycationic compound for the preparation of a medicament with
XX reduced inflammatory potential
XX
XX Example 1; Page 13; 43pp; English.
XX
XX The invention discloses the use of a polycationic compound for the
XX preparation of a medicament with reduced inflammatory potential, for
XX treating or preventing inflammation or for a vaccine containing an
XX antigen, possible acting as an adjuvant. The medicaments include
XX allergens, cytokines, chemokines, immunostimulatory nucleic acids,
XX cytotoxic or antioligogenic drugs and compounds needed for wound healing.
XX The medicament acts locally at the site of administration, and lowers or
XX completely eliminates inflammatory side effects of medicaments. Thus the
XX medicament reduces the inflammatory potential of a medicament and allows
XX the administration of medicaments that are usually not administered or
XX only rarely administered due to their inflammatory side effects. The
XX sequence presented is the OVA 257-264-peptide, a major histocompatibility
XX complex (MHC) class I (H-2Kb) restricted epitope of chicken ovalbumin
XX which was used in the scope of the invention.
XX
XX Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 SIINFEKL 8
DB 1 SIINFEKL 8
XX
XX RESULT 50
XX AAU99718
XX ID AAU99718 standard; Peptide; 8 AA.
XX
XX AC AAU99718;
XX
XX DT 07-OCT-2002 (first entry)
XX
XX DE Mouse MHC class I Kb OVA peptide sequence.
XX
XX KW Mutant major histocompatibility complex class I chimeric protein; MHC;
XX lymphocyte; T-cell receptor; tissue sample; biopsy material; pathogen;
XX bodily fluid; T lymphocyte; neoplastic cell; tumour cell; MHC antigen;
XX virus; protozoan; bacteria; fungi; nematode; immune response; activator;
XX enhancer; T cell activator; mouse; recombinant yeast cell; Kb; OVA; Id;
XX beta2m; dev8.
XX
XX OS Mus sp.
XX
XX PN WO200246399-A2.
XX
XX PD 13-JUN-2002.
XX
XX PE 10-DEC-2001; 2001WO-US47817.
XX
XX PR 08-DEC-2000; 2000US-254495P.
XX
XX PA (UNII ) UNIV ILLINOIS FOUNDD.
XX
XX PI Kranz DM, Brophy S;
XX
XX DR WPI; 2002-527916/56.
XX
XX

```

PT New isolated mutant major histocompatibility complex class I chimeric
PT protein displayed on surfaces of recombinant yeast cells, has improved
PT stability, and is useful for activating immune response -
XX
XX
XX Example 3; Figure 18; 96bp; English.
XX
XX The present invention relates to a new mutant major histocompatibility
CC complex (MHC) class I chimeric protein. The protein of the invention
CC comprises a portion mediating binding to surfaces of recombinant yeast
CC cells and a portion comprising peptide binding region of MHC class I
CC protein, where the invention is improved in stability as compared with
CC MHC class I chimeric protein which is not a mutant chimeric protein.
CC The protein, further comprising a detectable label, is useful for
CC detecting a lymphocyte having a T-cell receptor protein in a biological
CC sample such as cells, tissue sample, biopsy material or bodily fluids.
CC The method is useful for detecting a T lymphocyte that is specific for
CC a neoplastic cell, a tumour cell, a virus-infected cell, a protozoan-
CC infected cell, a bacterium-infected cell or a fungus-infected cell. The
CC protein of the invention can be used to directly activate T cells, in
CC order to identify/screen for peptide-MHC antigens. The protein is also
CC useful in activating T cells that participate in the removal of target
CC cells including neoplastic cells and cells infected with pathogenic
CC agents including viruses, protozoans, bacteria, fungi or nematodes.
CC The invention is improved in stability as compared with MHC class I
CC protein which is not a mutant chimeric protein. The present amino acid
CC sequence represents a mouse MHC peptide of the invention, as described
CC above.
XX
XX SQ Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEKL 8
XX |||||
XX 1 SIINFEKL 8
XX
XX DB 1 SIINFEKL 8
XX
XX RESULT 51
XX ABB08108
XX ID ABB08108 standard; peptide; 8 AA.
XX
XX AC ABB08108;
XX
XX DT 10-SEP-2002 (first entry)
XX
XX DE Chicken ovalbumin CTL epitope fragment.
XX
XX KW T cell; antigen; tumour; vaccine; cytostatic; cancer; ovalbumin;
XX CTL epitope.
XX
XX KM Gallus sp.
XX
XX OS
XX PN US6387701-B1.
XX
XX PD 14-MAY-2002.
XX
XX PF 30-APR-1999; 99US-0302329.
XX
XX PR 30-APR-1996; 96US-0640444.
XX PR 30-APR-1997; 97WO-US07317.
XX PR 06-MAY-1998; 98US-0073819.
XX PR 16-FEB-1999; 99US-0171916.
XX
XX PA (UYDU-) UNIV DUKE.
XX
XX PI Nair SK, Boczkowski DJ, Gilboa E;
XX
XX DR MPI, 2002-478447/51.
XX
XX PT Identifying tumor antigens that elicit T cell responses and which may
XX be used for vaccinating against cancers, e.g. melanomas, breast

PT cancers, prostate cancers, colon cancers, and ovarian cancers,
PT comprises a cytotoxicity assay -
XX
XX
XX PS Example 1; Column 12; 21pp; English.
XX
XX CC The invention relates to identifying a tumour antigen that elicits a T
CC cell response directed against the tumour. The antigen may then be used
CC to vaccinate against cancers e.g. melanomas, bladder cancers, breast
CC cancers, pancreatic cancers. The present sequence represents a chicken
CC ovalbumin CTL epitope fragment.
XX
XX SQ Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEKL 8
XX |||||
XX 1 SIINFEKL 8
XX
XX DB 1 SIINFEKL 8
XX
XX RESULT 52
XX ABB81273
XX ID ABB81273 standard; peptide; 8 AA.
XX
XX AC ABB81273;
XX
XX DT 20-AUG-2002 (first entry)
XX
XX DE Chicken OVA 257-264 peptide SEQ ID NO:1.
XX
XX KW Yeast; dendritic cell; vaccine; immune response; ovalbumin; antifungal;
XX immunostimulant; antibacterial; virucide; antiprotozoal; cytostatic;
XX immunisation; cell mediated immunity; infectious disease; cancer.
XX
XX OS Gallus gallus.
XX
XX PN WO200239951-A2.
XX
XX PD 23-MAY-2002.
XX
XX PF 15-NOV-2001; 2001WO-US43537.
XX
XX PR 15-NOV-2000; 2000US-249173P.
XX
XX PA (GLOBE-) GLOBE IMMUNE INC.
XX PA (UYTE-) UNIV TECHNOLOGY CORP.
XX
XX PI Duke RC, Bellgrau D, Franzusoff A, Wilson CC;
XX
XX DR MPI, 2002-479895/51.
XX
XX PT Therapeutic composition, useful as vaccine, comprises dendritic cell
XX intracellularly loaded with yeast vehicle and at least one antigen -
XX
XX PS Example 4; Page 34; 68pp; English.
XX
XX CC The present invention describes a therapeutic composition (I) comprising
CC a dendritic cell (II), a yeast vehicle (III) and at least one antigen
CC (IV), where (II) has been loaded intracellularly with (III) and (IV).
CC (I) has immunostimulant, antibacterial, antifungal, virucide,
CC antiprotozoal and cytostatic activities. (I) has many attributes that
CC make it an ideal vaccine candidate, including ease of construction, low
CC expense of mass production, biological stability and safety. No grossly
CC adverse side effects of immunisation with whole yeast were apparent at
CC the time of the initial vaccination or upon real administration. The
CC composition provides a powerful strategy for the induction of cell-
CC mediated immunity directed against a variety of infectious diseases and
CC cancer targets. The present sequence represents a chicken OVA (ovalbumin)
CC 257-264 peptide which is used in an example from the present invention.
XX
XX SQ Sequence 8 AA;
XX

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 53

AAE22531
 ID AAE22531 standard; peptide; 8 AA.

AC AAE22531;

DT 26-JUL-2002 (first entry)

DE Ovalbumin peptide to evaluate the ability of aroC/sifa (P3H8) TML mutant.

KW Sifa; attenuated microorganism; medicament; allergen; gene therapy;
 KW vaccine; protein therapy; virucide; hepatotropic; antiinflammatory;
 KW protozoecide; ovalbumin.

OS Unidentified.

PN MO200226251-A1.

PD 04-APR-2002.

PF 01-OCT-2001; 2001WO-GB04358.

PR 29-SEP-2000; 2000GB-0023906.

PR 14-AUG-2001; 2001GB-0019802.

PA (MIGR-) MICROSCIENCE LTD.

PI Brennan FR, Dougan G;

PS WPI; 2002-339986/37.

PT An attenuated Salmonella strain, for producing an elevated immune
 PT response treat diseases, comprises a mutation that disrupts expression
 PT of the sifa gene, and expresses a therapeutic heterologous peptide such
 PT as an antigen -

XX Example 1; Page 7; 24pp; English.

CC The present invention relates to an attenuated Salmonella microorganism
 CC which comprises a mutation that disrupts the expression of the sifa gene
 CC and expresses a therapeutic heterologous peptide such as an antigen. The
 CC attenuated microorganisms are useful for manufacturing medicaments to
 CC treat or prevent a disease which can be treated by the heterologous
 CC product and to increase the MHC class I-restricted response in a patient,
 CC to deliver a therapeutic polynucleotide to a host cell to treat a disease
 CC which can be corrected by administering the polynucleotide and to cause
 CC an increase in the MHC class I-restricted response in a patient. They
 CC are also useful to deliver heterologous antigens and allergens to a
 CC patient, such as hepatitis, herpes simplex and Malarial antigens. The
 CC method of the invention is useful for delivery of antisense nucleotides
 CC or ribozymes in gene therapy. Sequences of the invention are also used
 CC as vaccines and in protein therapy. The present sequence is ovalbumin
 CC peptide (OVA 257-264) used to evaluate the ability of an aroC/sifa (P3H8)
 CC TML mutant to stimulate a MHC class I-restricted response to heterologous
 CC antigen.

SO Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8

DB |||||
 1 SIINFEXL 8

RESULT 54

ABB76050
 ID ABB76050 standard; Peptide; 8 AA.

AC ABB76050;

DT 12-JUL-2002 (first entry)

DE Ovalbumin, H-2kb restricted epitope.

KW Bordetella pertussis; adenylate cyclase; CyA; adenylcyclase;
 KW vector; drug delivery; antigen delivery; cell targeting; CD11b;
 KW ovalbumin; epitope; chicken.

OS Gallus sp.

PN EP1188446-A1.

PD 20-MAR-2002.

PF 15-SEP-2000; 2000EP-0402562.

PR 15-SEP-2000; 2000EP-0402562.

PA (INSP) INST PASTEUR.

PA (CNRS) CENT NAT RECH SCI.

PI Lelerc C, Guernompres P, Ladant D, Guiso N, Khelef N;

PS WPI. 2002-354020/39.

PT Use of Bordetella adenylcyclase to make proteinaceous vector, useful
 PT for drug or antigen delivery, selectively targets cells that express
 PT CD11b -

XX Example B; Page 14; 34pp; English.

CC The present sequence is the peptide sequence of a chicken ovalbumin,
 CC H-2kb restricted epitope, which was used as an experimental model
 CC epitope in an example from the invention. The epitope was
 CC genetically inserted into the catalytic domain of a detoxified, but
 CC still invasive, mutant adenylate cyclase (adenylcyclase, CyA) of
 CC Bordetella pertussis. The recombinant toxin, CyAOVA, was used to
 CC immunise C57BL/6(H-2b) mice once i.v. CD4- and CD40-independent
 CC cytotoxic T lymphocyte (CTL) priming was observed in the absence
 CC of adjuvant. The invention relates to the novel use of Bordetella
 CC CyA as a proteinaceous vector for targeting a molecule of interest
 CC to the surface CD11b-expressing cells, especially dendritic cells
 CC and neutrophils. The molecule of interest is translocated in the
 CC cytosol to prime a CTL response. In a preferred embodiment, a
 CC peptide is inserted into the catalytic domain of CyA at a
 CC permissive site. The peptide may be an intracellular bacterial
 CC cell, tumour, viral, fungal or parasite cell antigen (all claimed).
 CC Alternatively, a drug, especially an antiinflammatory, is chemically
 CC coupled to CyA for drug delivery.

SO Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 55
 AAB19945

ID AAE19945 standard; peptide; 8 AA.
 XX AAE19945;
 AC
 XX
 DT 18-JUN-2002 (first entry)
 DE Cytotoxic T-cell epitope for ovalbumin.
 XX
 DE Cytotoxic T-cell; CTL; tumour; cancer; infection; cell-mediated immunity;
 KW vaccine; immune response; cytostatic; T-cell epitope; ovalbumin.
 KM
 XX
 OS Unidentified.
 XX
 PN US2002018785-A1.
 XX
 PD 14-FEB-2002.
 XX
 PF 02-APR-2001; 2001US-0822250.
 XX
 PR 22-SEP-1997; 97US-0935377.
 XX
 PA (UVRP) UNITV ROCHESTER.
 XX
 PI Zauderer M;
 XX
 DR WPI; 2002-239252/29.
 XX
 PT Representational Difference Analysis method for identification of
 XX antigens recognized by cytotoxic T cells and specific for human tumors,
 PT comprises improved selection of genes encoding target antigens -
 XX
 PS Example 1; Page 13; 54pp; English.
 XX
 CC The present invention relates to novel methods for the identification
 CC of antigens recognised by cytotoxic T cells (CTLs) and specific for
 CC human tumours, cancers and infected cells. The method involves screening
 CC the products of an expression library generated from DNA/RNA of a cell
 CC expressing a target epitope with cytotoxic T cells generated against
 CC the cell to identify DNA clones expressing target epitope or providing
 CC cytotoxic T cells specific for a gene product differentially expressed
 CC by a cell and measuring the cross-reactivity of the cytotoxic T cells
 CC for cells expressing a target epitope in which the target epitope is
 CC identified as a gene product inducing cytotoxic T cells. The method is
 CC useful for identifying a target epitope or antigen specific for a tumour
 CC cell. The target epitope is also useful for identifying target antigens
 CC in other target cells against which it is desirable to induce cell-
 CC mediated immunity. The antigen identified by the method is useful
 CC in immunogenic compositions and vaccine preparations to induce the
 CC regression of tumours, cancers and infections in mammals. The invention
 CC also relates to vaccinia viral vectors which are useful for treating
 CC tumour-bearing mammals, including humans to generate immune response
 CC against the tumour cells. They are also useful for immunising or
 CC vaccinating tumour-free subjects to prevent tumour formation. The
 CC present sequence is cytotoxic T-cell epitope for ovalbumin. This
 CC peptide is used in the exemplification of the invention.
 XX
 SQ Sequence 8 AA;
 XX
 QY Query Match 100.0%; Score 38; DB 23; Length 8;
 XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 DB Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX 1 STINPEKL 8
 XX 1 STINPEKL 8
 DB 1 STINPEKL 8
 XX
 RESULT 56
 ABB09907
 ID ABB09907 standard; peptide; 8 AA.
 XX
 AC ABB09907;
 XX

DT 10-JUN-2002 (first entry)
 XX
 DE Ovalbumin peptide (H-2kb CTL epitope).
 XX
 DE Ovalbumin; cytokine; ELISPOT assay; polycationic substance;
 KW cytokine secreting cell; interleukin; IL; interferon; IFN; TNF; CSF;
 KW tumour necrosis factor; colony stimulating factor;
 KW enzyme-linked immunosorbent spot assay.
 XX
 OS Aves.
 XX
 PN WO200179854-A1.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-EP04208.
 XX
 PR 13-APR-2000; 2000AT-0000645.
 XX
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Schallich J, Buschle M;
 XX
 DR WPI; 2002-240136/29.
 XX
 PT Identification of cytokine secreting cells involves incubating cell
 XX suspension containing antigen specific cells in the presence of
 PT polycationic substance -
 XX
 PS Example 1; Page 10; 22pp; English.
 XX
 CC The sequence represents an ovalbumin peptide (H-2kb CTL epitope from
 CC ovalbumin). The invention relates to a novel method for identifying
 CC cytokine secreting cells (especially ELISPOT assays) using a polycationic
 CC substance. The method is useful for identifying cytokine secreting cells.
 CC The cytokine may be selected from interleukins (IL), interferons (IFN),
 CC tumour necrosis factors (TNF), and colony stimulating factors (CSF).
 XX
 SQ Sequence 8 AA;
 XX
 QY Query Match 100.0%; Score 38; DB 23; Length 8;
 XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 DB Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX 1 STINPEKL 8
 XX 1 STINPEKL 8
 DB 1 STINPEKL 8
 XX
 RESULT 57
 AAU76942
 ID AAU76942 standard; Peptide; 8 AA.
 XX
 AC AAU76942;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE OVA peptide (257-264).
 XX
 DE ODN; immunostimulatory oligodeoxynucleotide; Ovalbumin; OVA;
 KW cardiant; vaccine; tuberculosis; diphteria; pertussis; measles;
 KW tetanus; acquired immune deficiency syndrome; AIDS; malaria;
 KW cardiovascular disease; cancer; deoxyinosine; chicken.
 XX
 OS Gallus sp.
 XX
 PN WO200193905-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 07-JUN-2001; 2001WO-EP06433.
 XX
 PR 08-JUN-2000; 2000AT-0001000.
 XX

PR 23-NOV-2000; 2000AT-0001973.
 XX
 XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Schmidt W, Lingnau K, Schellack C, Egyed A;
 XX
 XX WPI; 2002-240927/29.
 DR
 XX
 XX
 PT New oligodeoxynucleic acid molecule useful as an immunostimulatory
 agent -
 XX
 PS Example 1; Page 18; 52pp; English.
 XX
 XX This invention relates to immunostimulatory oligodeoxynucleic acid
 CC molecules (ODN) that can be used to enhance an immune response for
 CC use in vaccines. The immunostimulatory oligonucleotides of the invention
 CC and pharmaceutical compounds containing them may be used as medicine,
 CC especially as an immunostimulatory agent, for the preparation of
 CC vaccines useful for the treatment of tuberculosis, diphtheria,
 CC pertussis, measles and tetanus, acquired immune deficiency syndrome
 CC (AIDS), malaria, cardiovascular diseases, and cancer. Oligonucleotides
 CC containing deoxyinosine residues (I-ODN) show a better immunostimulatory
 CC effect compared to prior art compounds containing CPG motifs. The ODNs
 CC of the invention produce more specific immune response to a given
 CC antigen or antigen fragment than the prior art compounds containing Cpg.
 CC Using immunostimulatory oligonucleotides containing deoxyinosine reduces
 CC the induction of adverse side reactions, especially the induction of
 CC systemic TNF-alpha or interleukin-6. The immunostimulatory effect of the
 CC composition containing a polycationic polymer and an antigenic fragment
 CC was significantly higher than could be expected from the addition of the
 CC effects of each single component or even the addition of the effects of
 CC the ODN or the polycation with the antigen. The present sequence
 CC represents the ovalbumin (OVA) peptide used as an antigen in
 CC examples of the method of the invention.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8
 RESULT 58
 AAU76802
 ID AAU76802 standard; Peptide; 8 AA.
 AC
 XX AAU76802;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE MHC class I-restricted epitope of chicken ovalbumin.
 XX
 KM Chicken; ovalbumin; immunostimulant; T cell epitope; inosine; cytosine;
 KM polycationic peptide; systemic immune response; MHC class I; vaccine;
 KM major histocompatibility complex class I.
 XX
 OS Gallus gallus.
 XX
 PN WO200193903-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 07-JUN-2001; 2001WO-EP06437.
 XX
 PR 08-JUN-2000; 2000AT-0001000.
 XX
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Egyed A, Lingnau K, Matzner F, Buschle M, Schmidt W;

XX
 DR WPI; 2002-205813/26.
 XX
 XX Pharmaceutical composition for the preparation of vaccine comprises T
 PT cell epitope(s) or its mixture, polycationic peptide and nucleic acid
 PT based on inosine and cytosine -
 XX
 XX
 PS Example 1; Page 10; 45pp; English.
 XX
 XX The invention relates to a pharmaceutical composition comprising a T cell
 CC epitope(s) or its mixture, a polycationic peptide and a nucleic acid
 CC based on inosine and cytosine. The composition of the invention induces a
 CC systemic immune response and is used for the preparation of a vaccine.
 CC This sequence represents an MHC class I-restricted epitope of chicken
 CC ovalbumin, used to test enhancement of immune response against an
 CC ovalbumin-derived peptide.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8
 RESULT 59
 AAU76869
 ID AAU76869 standard; Peptide; 8 AA.
 AC
 XX AAU76869;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE OVA peptide fragment.
 XX
 KM Membrane vesicle; exosome; density cushion centrifugation;
 KM dendritic cell; MHC; major histocompatibility complex; CDI; tumour;
 KM immunotherapy treatment; cancer; infection; immune disease; antitumour;
 KM cytotatic; OVA.
 XX
 OS Unidentified.
 XX
 PN WO200182958-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 11-APR-2001; 2001WO-EP04173.
 XX
 PR 27-APR-2000; 2000US-0561205.
 PR 09-FEB-2001; 2001US-0780748.
 XX
 PA (APCE-) AP CELLS INC.
 XX
 PI Lamparski H, Ruegg C, Le Pecq J, Hsu D, Yao J;
 XX
 DR WPI; 2002-066489/09.
 XX
 PT Preparing membrane vesicle from biological sample for treating cancer,
 PT by culturing membrane vesicle-producing cells to release vesicles,
 PT enriching vesicles and subjecting sample to density cushion
 PT centrifugation -
 XX
 XX
 PS Example 12; Page 55; 103pp; English.
 XX
 XX The invention relates to a method for preparing membrane vesicles (in
 CC particular exosomes) from a biological sample, comprising culturing a
 CC population of membrane vesicle-producing cells under conditions allowing
 CC the release of the vesicles, enriching the vesicles and treating the
 CC enriched biological sample by density cushion centrifugation. Immunogenic
 CC membrane vesicles are useful for producing an immune response in a

CC subject, by obtaining a biological sample containing dendritic cells,
 CC isolating or purifying a membrane vesicle from the sample, contacting the
 CC purified vesicle with a peptide or a lipid under conditions allowing the
 CC peptide or lipid to bind an MHC or CD1 molecule at the surface of the
 CC vesicle, and administering the vesicle to the subject. Membrane vesicles
 CC are useful for immunotherapy treatment or prophylaxis of tumours, and for
 CC treating various disease conditions such as cancer, infections, and
 CC immune diseases. This sequence represents an OVA peptide fragment which
 CC is contacted with an exosome, to create an immunogenic membrane vesicle.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 60

AAU11866
 ID AAU11866 standard; peptide; 8 AA.

XX AAU11866;

DT 26-MAR-2002 (first entry)

DE Ovalbumin derived control peptide.

XX T0 terminator; PSA; DNA vaccine; anti-HIV; virucide;
 XX Human Immunodeficiency Virus; HIV; Gag; HIV gp120; HIV Pol; HIV Env;
 XX HIV VLP; measles fusion protein; measles haemagglutinin; epitope;
 XX measles nucleoprotein; influenza haemagglutinin; C3d gene; ovalbumin;
 XX cell-mediated immune response; humoral immune response; infection.

OS Unidentified.

PN WO200192470-A2.

PD 06-DEC-2001.

PF 02-MAR-2001; 2001WO-US06795.

XX 02-MAR-2000; 2000US-186364P.

PR 01-DEC-2000; 2000US-251083P.

XX (UYEM-) UNIV EMORY.

PI Robinson HL, Smith JM, Ross TM, Bright RA, Hua J, Ellenberger D;

XX WPI; 2002-075465/10.

DR Novel pGA vector useful for immunising patient against measles,

PT influenza has termination sequence encoding lambda T0 terminator and a

PT eukaryotic transcription cassette with vaccine insert encoding

XX immunogens of pathogens -

XX Example 14; Page 61; 174pp; English.

XX The invention relates to a vector (a pGA construct) comprising a
 CC termination sequence coding for the lambda T0 terminator, a prokaryotic
 CC origin of replication, a selectable marker gene and a eukaryotic
 CC transcription cassette comprising a vaccine insert encoding one or more
 CC immunogens derived from a pathogen e.g. Human Immunodeficiency Virus
 CC (HIV) Gag, HIV gp120, HIV Pol, HIV Env, HIV VLP, or its mutants, measles
 CC fusion protein, measles haemagglutinin, measles nucleoprotein, influenza
 CC haemagglutinin, or its mutants, or subsequences, and optionally, at least
 CC one C3d gene, is useful for immunising or treating a patient, when
 CC administered by an intramuscular or intradermal route. The immunisation
 CC methods using pGA elicit both cell-mediated and humoral immune responses
 CC that may limit the infection, spread or growth of the pathogen and result

CC in protection against subsequent challenge against the pathogen. The
 CC terminator sequence present prevents read-through from the kanamycin
 CC cassette into vaccine sequences while the plasmid is being produced in
 CC bacteria. Prevention of transcriptional read-through stabilises vaccine
 CC insert sequences by limiting the exposure of secondary structures that
 CC can be recognised by bacterial endonucleases. The present sequence
 CC is an ovalbumin control peptide used in an experiment to measure the
 CC T-cell response in monkeys inoculated with a pGA vector carrying vaccinia
 CC virus genes.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 61

AAU11239
 ID AAU11239 standard; peptide; 8 AA.

XX AAU11239;

DT 12-MAR-2002 (first entry)

DE Immunodominant Kb-restricted Cytotoxic T lymphocyte epitope #2.

XX Cyostatic; vaccine; tetanus toxin; Frc; tumour; CTL;
 XX cytotoxic T-lymphocyte; immunodominant Kb-restricted CTL epitope.

OS Unidentified.

PN WO200179510-A1.

XX 25-OCT-2001.

PF 17-APR-2001; 2001WO-GB01719.

XX 17-APR-2000; 2000GB-0009470.

XX (CANC-) CANCER RES VENTURES LTD.

PI Rice J, Stevenson F;

XX WPI; 2002-066370/09.

DR Nucleic acid construct, useful to immunise against various diseases

PT including cancer, expresses the first domain of tetanus toxin Frc fused

PT to a disease peptide antigen to provide a vaccine -

XX Disclosure; Page 25; 71pp; English.

XX The invention relates to a nucleic acid construct for delivery into
 CC living cells in vivo, to induce an immune response to a disease peptide
 CC antigen, where the construct directs expression of a fusion protein
 CC comprising the peptide antigen and the first domain of Frc. Also
 CC included are a nucleic acid vector comprising the above construct,
 CC a host cell comprising the above construct or vector and a method of
 CC producing a nucleic acid construct for inducing an immune response.
 CC The method comprises identifying a nucleic acid sequence encoding a
 CC disease peptide antigen comprising epitopes characteristic of the
 CC disease, cloning the nucleic acid sequence, introducing the cloned
 CC nucleic acid into a vector which allows the antigen to be expressed as a
 CC fusion with a first domain Frc from tetanus toxin, and optionally
 CC isolating the construct from the vector. The construct or vector is used
 CC as a vaccine to induce an immune response, particularly to tumour
 CC antigens. The present sequence is an immunodominant Kb-restricted
 CC cytotoxic T-lymphocyte (CTL) epitope suitable for inclusion in the
 CC vaccine of the invention.

```

XX  SQ      Sequence      8 AA;
XX  Query Match      100.0%; Score 38; DB 23; Length 8;
XX  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX  Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      1 SIINFEKL 8
      |||||
Db      1 SIINFEKL 8

RESULT 62
AAU09820
ID      AAU09820 standard; peptide; 8 AA.
XX
XX  AAU09820;
XX
XX  14-FEB-2002 (first entry)
XX
XX  Ovalbumin-derived class I H-2Kb restricted peptide (245).
XX
XX  Ovalbumin-derived class I H-2Kb restricted peptide (245); vaccine;
XX  immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
XX  popliteal lymph node; spleen; immune response; systemic response.
XX
XX  Unidentified.
XX
XX  WO200178767-A2.
XX
XX  25-OCT-2001.
XX
XX  17-APR-2001; 2001WO-EP04313.
XX
XX  14-APR-2000; 2000AT-0000657.
XX
XX  (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX  Matner F, Zauner W, Schmidt W, Buschle M;
XX
XX  WPI; 2002-025970/03.
XX
XX  Pharmaceutical preparation for use as a potent vaccine for inducing an
XX  improved immune response in a mammal, comprises a modified peptide -
XX
XX  Example 1; Page 8; 18pp; English.
XX
XX  The invention relates to a pharmaceutical preparation comprising a
XX  modified peptide, which induces an improved immune response in a mammal
XX  compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
XX  negatively charged (Glu), one positively charged (Iys) amino acid) was
XX  rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
XX  Asp, respectively. Results showed that the addition of 4 negatively-
XX  charged amino acids (EDBD) at the N-terminus of peptide SIINFEKL makes
XX  this peptide (in combination with poly-L-arginine) able to induce a high
XX  amount of specific interferon (IFN)-gamma-producing T cells in the
XX  draining (popliteal) lymph node (local response) and in the spleen
XX  (systemic response). Thus, the addition of hydrophobic amino acids as
XX  well as the addition of negatively charged amino acids transforms the
XX  peptide SIINFEKL to a good inducer of specific T cells. The modified
XX  peptides of the pharmaceutical composition induce a stronger immune
XX  response in a mammal compared to wild type antigens. The present
XX  sequence represents ovalbumin-derived class I H-2Kb restricted peptide
XX  (245) used to produce the modified hydrophobic peptides described in
XX  the method of the invention.
XX
SQ      Sequence      8 AA;
XX
XX  Query Match      100.0%; Score 38; DB 23; Length 8;
XX  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX  Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
Qy      1 SIINFEKL 8

```

```

Db      1 SIINFEKL 8
      |||||
XX
XX  RESULT 63
XX  AAE13215
XX  AAE13215 standard; peptide; 8 AA.
XX
XX  AAE13215;
XX
XX  12-FEB-2002 (first entry)
XX
XX  Hen egg ovalbumin (OVA) peptide.
XX
XX  Cytotoxic T lymphocyte; CTL; T cell; tumour load; cancer radiotherapy;
XX  immunostimulatory sequence oligonucleotide; ISS-ODN; chemotherapy;
XX  immunosuppression; transplantation; autoimmune disease; infection;
XX  acquired immune deficiency syndrome; AIDS; intracellular pathogen;
XX  cytomegalovirus; mycobacterial infection; Epstein-Barr virus;
XX  varicella zoster virus; human immunodeficiency virus; HIV; hen;
XX  ovalbumin; OVA.
XX
XX  Unidentified.
XX
XX  WO200172123-A1.
XX
XX  04-OCT-2001.
XX
XX  28-MAR-2001; 2001WO-US10118.
XX
XX  28-MAR-2000; 2000US-192537P.
XX
XX  11-MAY-2000; 2000US-203567P.
XX
XX  05-JUL-2000; 2000US-215895P.
XX
XX  (REGC) UNIV CALIFORNIA.
XX  (VETE-) DEPT VETERANS AFFAIRS.
XX
XX  Raz E, Cho HJ, Richman DD, Horner AA;
XX
XX  WPI; 2002-010699/01.
XX
XX  Increasing antigen-specific cytotoxic T lymphocyte activity in a CD4+ T
XX  cell deficient individual, useful to treat immunodeficiency and block
XX  HIV infection, comprises administering immunostimulatory nucleic acid
XX
XX  Example 1; Page 44; 91pp; English.
XX
XX  The present invention relates to a method for increasing antigen-specific
XX  cytotoxic T lymphocyte (CTL) activity in a CD4+ T cell-deficient
XX  individual, comprising administering an immunostimulatory sequence
XX  oligonucleotide (ISS-ODN). The immunostimulatory nucleic acids of the
XX  invention are used in CD4+ T cell-deficient individuals to decrease
XX  tumour load, to treat a primary or acquired immunodeficiency,
XX  particularly where the acquired immunodeficiency is temporary and due
XX  to cancer radiotherapy or chemotherapy or immunosuppression following
XX  bone marrow or organ transplantation, or autoimmune disease treatment,
XX  or is acquired immunodeficiency syndrome (AIDS). The nucleic acids may
XX  be used to treat a person at risk of becoming CD4+ T cell-deficient,
XX  particularly where someone at risk of cancer recurrence. They are also
XX  used to treat infection, particularly by an intracellular pathogen,
XX  especially one caused by cytomegalovirus, Mycobacterium tuberculosis,
XX  M. avium, Epstein-Barr virus, a fungus yeast, varicella zoster virus or
XX  human immunodeficiency virus (HIV). The present sequence is a hen egg
XX  ovalbumin (OVA) peptide, used in the exemplification of the invention.
XX
SQ      Sequence      8 AA;
XX
XX  Query Match      100.0%; Score 38; DB 23; Length 8;
XX  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX  Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
Qy      1 SIINFEKL 8

```

```

Db      1 STINFEKL 8
      |||||
RESULT 64
AAE13436
ID      AAE13436 standard; peptide; 8 AA.
XX
AC      AAE13436;
XX
DT      12-FEB-2002 (first entry)
XX
DE      Chicken ovalbumin major histocompatibility complex class I epitope.
XX
KW      Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW      major histocompatibility complex; MHC; therapy; immune response;
KW      malignancy; chicken.
XX
OS      Gallus gallus.
XX
PN      WO200179259-A1.
XX
PD      25-OCT-2001.
XX
PF      17-APR-2001; 2001WO-US12567.
XX
PR      17-APR-2000; 2000US-197462P.
XX
PA      (ROTH/) ROTHMAN J E.
PA      (MAYH/) MAYHEW M.
PA      (HOEW/) HOE M.
XX
PI      Rothman JE, Mayhew M, Hoe M;
XX
DR      WPI; 2002-017594/02.
XX
PT      A new antigenic complex comprising epitopes non-covalently joined to a
PT      heat shock protein by a molecular tether designated a javelin are
PT      useful to treat or prevent infectious disease or malignancy -
XX
XX      Example; Page 15; 47pp; English.
XX
PS      The present invention relates to an antigenic complex, comprising a
CC      number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC      a tethering molecule referred to as javelin which has affinity for the
CC      HSP under physiological conditions, where the epitopes are covalently
CC      joined to the tethering molecule and one epitope is major
CC      histocompatibility complex class I (MHC) and the other MHC class II. The
CC      antigenic complex is used to induce immune responses directed towards the
CC      treatment or prevention of infectious diseases and malignancies. The
CC      present sequence is chicken ovalbumin MHC class I epitope.
XX
SQ      Sequence 8 AA;
XX
Query Match      100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 STINFEKL 8
      |||||
DB      1 STINFEKL 8
XX
RESULT 65
AAU75056
ID      AAU75056 standard; peptide; 8 AA.
XX
AC      AAU75056;
XX
DT      23-APR-2002 (first entry)
XX
DE      Ovalbumin antigenic peptide.
XX

```

```

KW      Ovalbumin; Th1; Th2; antigen; immunosuppressive; cytostatic;
KW      adjuvant; lipopolysaccharide; lipid; vaccine; immunogenicity;
KW      immunocompetence; autoimmune disease; infectious disease; OVA;
KW      graft-versus- host disease; tumour; transgenic T cell; chicken.
XX
OS      Gallus gallus.
XX
PN      WO200197838-A2.
XX
PD      27-DEC-2001.
XX
PF      18-JUN-2001; 2001WO-US19411.
XX
PR      16-JUN-2000; 2000US-212182P.
XX
PA      (BAYU ) BAYLOR RES INST.
XX
PI      Pulendran B, Bancherreau JF, Cutler CW;
XX
DR      WPI; 2002-114543/15.
XX
PT      Use of adjuvants comprising isolated lipid groups such as Porphyromonas
PT      gingivalis lipopolysaccharides or its detoxified forms or derivatives
PT      for preparation of compositions to elicit T-helper cell responses
PT      in mammals -
XX
PS      Disclosure; Page 20; 58pp; English.
XX
CC      This invention relates to the use of adjuvants comprising isolated lipid
CC      groups such as Porphyromonas gingivalis lipopolysaccharide, its
CC      detoxified forms or derivatives, for preparing compositions for
CC      eliciting Th2 responses, enhancing vaccine immunogenicity, modulating
CC      immunocompetence, treating autoimmune/infectious disease, stimulating
CC      interleukin-5 (IL-5)/IL-13 production or dampening interferon gamma
CC      production in mammal. The adjuvant or a pharmaceutical compound
CC      containing it is useful for enhancing antibody harvest in a laboratory
CC      animal through an elicited Th2 immune response, or by modulating Th2
CC      immune responses. The adjuvant is also useful to study the Th2
CC      immune response in laboratory animal research, for the treatment or
CC      prophylactic vaccination of humans or animals against graft-versus-
CC      host disease, and for treating tumours. The present sequence
CC      represents the Ovalbumin (OVA) peptide used as an antigen to
CC      stimulate transgenic T cells in the method of the invention.
XX
SQ      Sequence 8 AA;
XX
Query Match      100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 STINFEKL 8
      |||||
DB      1 STINFEKL 8
XX
RESULT 66
ABU07743
ID      ABU07743 standard; Peptide; 8 AA.
XX
AC      ABU07743;
XX
DT      23-MAY-2003 (first entry)
XX
DE      Chicken ovalbumin epitope presented on murine Kb MHC class I.
XX
KW      Chicken; ovalbumin; OVA; cytolysin; vaccine delivery; Kb MHC class I;
KW      intracellular delivery vehicle; nonvirulent bacterium; drug delivery;
KW      gene therapy; biosynthesis; high level protein delivery;
KW      major histocompatibility complex; cytosolic protein delivery.
XX
OS      Gallus gallus.
XX
PN      US2002142007-A1.
XX

```

```

XX 03-OCT-2002.
PD 07-SEP-2001; 2001US-0949109.
XX
XX 13-AUG-1998; 98US-0133914.
PF 21-DEC-1999; 99US-0469197.
XX
XX (PORT/) PORTNOY D A.
XX (HIGGS/) HIGGINS D E.
XX
XX Portnoy DA, Higgins DE;
XX
XX WPI; 2003-328328/31.
XX
XX New nonvirulent bacterium with genes coding for a non-secreted foreign
PT cytolysin or a different foreign agent, useful as an intracellular
PT delivery vehicle for delivering, e.g. vaccines, drugs or genes for
PT therapy to eukaryotic cells
XX
XX Example; Page 6; 14pp; English.
XX
XX The invention relates to a nonvirulent bacterium, which comprises a first
CC gene encoding a non-secreted foreign cytolysin operably linked to a
CC heterologous promoter and a second gene encoding a different foreign
CC agent. The nonvirulent bacterium is useful as an intracellular delivery
CC vehicle, particularly of agents to eukaryotic cells. The nonvirulent
CC bacterium is particularly useful for delivering foreign agents for
CC diagnosis, therapy (e.g. prophylaxis such as vaccine, delivery of
CC therapeutic drug, or gene therapy) or biosynthesis. The nonvirulent
CC bacterium is also useful for delivering nucleic acids that provide
CC templates for transcription or translation, or provide modulators of
CC transcription and/or translation. No protein purification is required
CC compared to prior art delivery systems. In addition, high levels of
CC protein can be delivered to the cytosol of virtually any cell and the
CC levels can be controlled through the use of inducible promoters. L.
CC monocytogenes LfO (listeriolysin or cytolysin) was transformed in
CC E.coli cells and used as a system to deliver chicken ovalbumin to the
CC cytosol of macrophages. The present sequence represents the chicken
CC ovalbumin epitope presented on murine Kb major histocompatibility
CC complex, MHC, class I.
XX
XX SQ Sequence 8 AA;
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 38; DB 24; Length 8;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEXL 8
XX |||||
XX 1 SIINFEXL 8
XX
XX Db
XX
XX RESULT 67
XX ABU08619
XX ID ABU08619 standard; Peptide; 8 AA.
XX
XX AC ABU08619;
XX
XX 23-MAY-2003 (first entry)
XX
XX XX Ovalbumin (OVA) residues 257-264.
XX
XX DE Cancer; tumour; antigen-presenting cell; APC; tumour cell conjugate;
XX cytokine; interleukin; interferon; IFN alpha, IFNbeta; IFNgamma;
XX tumour necrosis factor; TNF; transforming growth factor; TGF;
XX granulocyte-macrophage colony stimulating factor; GM-CSF; vaccine;
XX melanoma; kidney cancer; pulmonary carcinoma; hepatic carcinoma;
XX mammary cancer; prostatic carcinoma; gastric carcinoma; leukaemia;
XX ovalbumin-specific tetramer; (OVA)-specific tetramer; ovalbumin;
XX OVA.
XX
XX OS Gallus gallus.
XX
XX

```

```

XX US2002182194-A1.
XX
XX PD 05-DEC-2002.
XX
XX 25-MAR-2002; 2002US-0106173.
XX
XX 04-APR-2001; 2001CN-0105852.
XX
XX (SHAN-) SHANGHAI BRILLIANCE BIOTECH INST.
XX
XX Ju D, Tao Q, Ye D;
XX
XX WPI; 2003-328591/31.
XX
XX New antigen-presenting cell and tumor cell conjugates, where the
PT antigen-presenting cell is modified by a cytokine gene, useful for the
PT preparation of a medicine for the therapy of cancer or a vaccine for
PT the prophylaxis of cancer
XX
XX Example 1; Page 4; 23pp; English.
XX
XX The invention describes an antigen-presenting cell (APC)/tumour cell
CC conjugate, where the APC is modified by a cytokine gene (interleukin
CC IL-2, IL-3, IL-4, IL-6, IL-12, IL-18, interferon (IFN) alpha, IFNbeta,
CC IFNgamma, tumour necrosis factor (TNF), transforming growth factor (TGF)
CC and/or granulocyte-macrophage colony stimulating factor (GM-CSF)). The
CC antigen-presenting cell/tumour cell conjugate is useful for the
CC preparation of a medicine for the therapy of cancer or a vaccine for the
CC prophylaxis of cancer. The cancer includes melanoma, kidney cancer,
CC pulmonary carcinoma, hepatic carcinoma, mammary cancer, prostatic
CC carcinoma, gastric carcinoma and leukaemia. This is the amino acid
CC sequence of Ovalbumin (OVA) residues 257-264 used in the creation of
CC Ovalbumin (OVA)-specific tetramers used in the vaccine of the invention.
XX
XX SQ Sequence 8 AA;
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 38; DB 24; Length 8;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEXL 8
XX |||||
XX 1 SIINFEXL 8
XX
XX Db
XX
XX RESULT 68
XX ABP57401
XX ID ABP57401 standard; peptide; 8 AA.
XX
XX AC ABP57401;
XX
XX 23-APR-2003 (first entry)
XX
XX DE Synthetic dimer peptide.
XX
XX XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
XX OS Synthetic.
XX
XX PN WO2003000899-A1.
XX
XX 03-JAN-2003.
XX
XX 20-JUN-2002; 2002WO-GB02829.
XX
XX 22-JUN-2001; 2001GB-0015382.
XX
XX (UYBR-) UNIV BRISTOL.
XX
XX PI Hirst TR;
XX

```

```

XX DR WPI; 2003-175291/17.
XX
XX PT Use of a mutant form of B subunit of Escherichia coli heat labile
XX PT enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX PT target cell for treating viral infection or cancer
XX PS Example 5; Page 45; 84pp; English.
XX
XX CC The present invention describes a mutant form of B subunit of Escherichia
XX CC coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX CC from Vibrio cholerae which is useful for delivering an agent to a target
XX CC cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX CC immunogenic and immunomodulatory activity relative to the wild-type form
XX CC of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX CC a subject; (2) delivering the agent using the mutant to a target cell;
XX CC (3) a composition; and (4) a kit for delivering the agent to a target
XX CC cell. Mutant EtxB and CtxB have virulence and cytostatic activities and
XX CC can be used in vaccines. The mutant can be used for the preparation of
XX CC a medicament for delivering an exogenous peptide, which is the agent,
XX CC into the major histocompatibility complex (MHC) Class I antigen
XX CC processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX CC (CTL) response, or for separate, simultaneous or combined use for
XX CC treating viral infection or cancer. The mutant form of EtxB or CtxB
XX CC enters mammalian cells without inducing a potent anti-B-subunit response
XX CC and immunomodulatory response. It may be linked with an agent to
XX CC upregulate the presentation of the antigen or antigenic determinant.
XX CC The present sequence represents a peptide which is used in an example
XX CC from the present invention.
XX
XX SQ Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 24; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SIINFEXL 8
XX |||||
XX 1 SIINFEXL 8
XX
XX DB
XX
XX RESULT 69
XX ABP58359
XX ID ABP58359 standard; Peptide; 8 AA.
XX
XX AC ABP58359;
XX
XX DT 07-APR-2003 (first entry)
XX
XX DE Ovalbumin-derived peptide OVA257-264.
XX
XX KW Chicken; ovalbumin; allergen; immunostimulant;
XX KW oligodeoxynucleic acid; ODN; vaccine.
XX
XX OS Gallus sp.
XX
XX OS WO200295027-A2.
XX
XX PD 28-NOV-2002.
XX
XX PF 17-MAY-2002; 2002WO-EP05448.
XX
XX PR 21-MAY-2001; 2001AT-0000805.
XX
XX PA (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX PI Lingnau K, Schellack C, Schmidt W;
XX
XX DR WPI; 2003-183880/18.
XX
XX PT New oligodeoxynucleic acid molecules useful for the preparation of
XX PT vaccine -

```

```

XX PS Example 3; Page 23; 57pp; English.
XX
XX CC The present sequence is that of OVA257-264, a major
XX CC histocompatibility complex class I (H-2Kb)-restricted epitope of
XX CC chicken ovalbumin. An example from the invention describes the
XX CC generation of specific immune responses against this allergen-derived
XX CC peptide using deoxyridine monophosphate-modified oligonucleotide
XX CC U-ODN 13 (see AB224776). U-ODN 13 is an example of new
XX CC oligodeoxynucleic acid (ODN) molecules useful in the preparation of
XX CC vaccines. The invention is based on the discovery that ODNs
XX CC containing deoxyridine residues have an immunostimulatory effect
XX CC comparable to, or greater than, ODNs containing CpG motifs.
XX CC Combining the ODN with an antigen strongly increases the potential
XX CC of the antigen to raise the protection/immune response of a
XX CC vaccinated individual.
XX
XX SQ Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 24; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SIINFEXL 8
XX |||||
XX 1 SIINFEXL 8
XX
XX DB
XX
XX RESULT 70
XX ABG73081
XX ID ABG73081 standard; Peptide; 8 AA.
XX
XX AC ABG73081;
XX
XX DT 02-APR-2003 (first entry)
XX
XX DE MHC Class I peptide OVA-8.
XX
XX KW Antigen-specific T lymphocyte; MHC-antigen complex; MHC Class I peptide;
XX KW major histocompatibility complex; tumour-specific killer T cell;
XX KW virus-specific killer T cell; cytostatic; virulence; OVA-8.
XX
XX OS Synthetic.
XX
XX OS US2002151690-A1.
XX
XX PD 17-OCT-2002.
XX
XX PF 05-NOV-1999; 99US-0434965.
XX
XX PR 12-AUG-1997; 97US-0909549.
XX
XX PA (LUXE/) LUXEMBURG A T.
XX PA (JACK/) JACKSON M R.
XX PA (PETE/) PETER P A.
XX
XX PI Luxembourg AT, Jackson MR, Peter PA;
XX
XX DR WPI; 2003-182532/18.
XX
XX PT Enriching antigen-specific T lymphocytes, for purifying or expanding in
XX PT vitro tumour- or virus-specific killer T cells for cell therapy,
XX PT comprises capture of the lymphocytes on a substrate coated with
XX PT antigenic peptide-MHC complexes -
XX
XX PS Example 2; Page 5; 40pp; English.
XX
XX CC The invention relates to a method for enriching antigen-specific T
XX CC lymphocytes, comprising contacting a heterogeneous population of
XX CC antigen-specific T lymphocytes with a matrix comprising MHC-antigen
XX CC complexes for a period of time sufficient to allow the antigen-specific T
XX CC lymphocytes to interact with the matrix, and eluting the antigen-specific
XX CC T lymphocytes from the matrix to provide an enriched population of

```


CC antigen-specific T lymphocytes. The MHC-antigen complexes comprise one or
 CC more antigens. Also claimed is a matrix for capturing antigen-specific T
 CC lymphocytes, comprising a support having on its surface an immobilised
 CC Class I peptide and a predetermined amount of an antigen, or for
 CC capturing antigens, comprising a support having on its surface an
 CC immobilised empty Class I peptide which is capable of binding one or more
 CC antigens, and isolating antigen-specific T lymphocytes from a
 CC heterogeneous population of cells from a patient. The methods are useful
 CC for enriching antigen-specific T lymphocytes to purify and expand in
 CC vitro tumour and virus-specific killer T cells for cell therapy. The
 CC methods are also useful for isolating or preparing a population of
 CC antigen-specific T lymphocytes from a patient for treatment of the
 CC patient's disease or condition. This sequence represents an MHC Class I
 CC peptide used in the method of the invention.

CC Sequence 8 AA;

Query Match 100.0%; Score 38; DB 24; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 71

ABP56760
 ID ABP56760 standard; peptide; 8 AA.

AC ABP56760;

DT 31-MAR-2003 (first entry)

DE Ovalbumin derived peptide OVA 257-264.

XX Stabilisation; polycationic polymer; medicine; vaccination;

KM gene therapy; drug; ovalbumin.

OS Gallus sp.

OS Synthetic.

PN WO200294845-A2.

PD 28-NOV-2002.

PF 17-MAY-2002; 2002WO-EP05447.

PR 21-MAY-2001; 2001AT-0000805.

PA (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.

PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PI Schellack C, Lingnau K, Schmidt W;

DR WPI; 2003-140356/13.

XX Use of polycationic polymer for stabilization of nucleic acids -

PT Example 1; Page 6; 28pp; English.

PS The present invention describes a method for the stabilisation of nucleic

CC acids involving contacting nucleic acids with a polycationic polymer in

CC aqueous solution or suspension. The method can be used for the

CC stabilisation of nucleic acid which can be used in medicines e.g.

CC or gene therapy drug. The present sequence represents an ovalbumin

CC derived peptide, designated OVA 257-264, which is used in an example

CC from the present invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 24; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 72

ABP60027
 ID ABP60027 standard; Peptide; 8 AA.

AC ABP60027;

DT 07-MAR-2003 (first entry)

DE Ovalbumin antigenic peptide.

XX TOP; thimet oligopeptidase; EC3.4.25.15; cytostatic; tumour;

KM immunostimulant; major histocompatibility complex class I; MHC;

KM T-cell immunity; chicken.

OS Gallus gallus.

PN WO200279388-A2.

PD 10-OCT-2002.

PF 01-APR-2002; 2002WO-US10385.

PR 30-MAR-2001; 2001US-280669P.

PA (UTMA-) UNIV MASSACHUSETTS.

PI Rock KL, Goldberg AL;

DR WPI; 2003-103265/09.

XX New recombinant cell comprising an exogenously derived nucleic acid

PT coding for a thimet oligopeptidase polypeptide, useful for modulating

PT an antigenic response in a mammal for treating e.g., tumour -

PS Example 1; Page 50; 73pp; English.

CC The invention relates to a new recombinant cell comprising an exogenously

CC derived nucleic acid that codes for a thimet oligopeptidase (TOP)

CC polypeptide. The TOP polypeptide is overexpressed in the cell compared to

CC a wild-type cell from which the recombinant cell is derived. The activity

CC of TOP may be described as cyostatic and immunostimulatory. Thimet

CC oligopeptidase (TOP; EC3.4.25.15) plays a key role in modulating levels

CC of major histocompatibility complex (MHC) class I-presented peptides. The

CC recombinant host cell of the invention is useful for modulating an

CC antigenic response in a mammal. Methods of the invention are useful for

CC screening a test compound for its ability to serve as an immunomodulatory

CC agent and identifying an antigen resistant to thimet oligopeptidase

CC degradation. A method of the invention is useful for increasing CD8

CC T-cell immunity, which uses vaccination with a TOP inhibitor for

CC decreasing TOP expression or activity. The vaccination method uses

CC treated tumour cells, antigen bearing/pulsed dendritic cells or injection

CC of a viral vector. The recombinant host cell is useful for treating

CC tumours. The current sequence represents an ovalbumin antigenic peptide

CC that is used in an example from the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 24; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

```

RESULT 73
ABU1029
ID ABU1029 standard; Peptide; 8 AA.
XX
XX
AC ABU1029;
XX
XX
DT 04-FEB-2003 (first entry)
XX
DE Ovalbumin immunodominant epitope.
XX
XX
KW Epitope; antigen-specific immunoglobulin; Ig; early/late promoter;
KW heavy chain constant region; light chain constant region;
KW variable region; camelised Ig heavy chain variable region; MHC; CTL;
KW major histocompatibility class; cytotoxic T-lymphocyte.
XX
XX
OS Unidentified.
XX
XX
PN US2002123057-A1.
XX
XX
PD 05-SEP-2002.
XX
XX
PF 14-NOV-2001; 2001US-0987456.
XX
XX
PR 17-NOV-2000; 2000US-249268P.
PR 18-JAN-2001; 2001US-262067P.
PR 27-FEB-2001; 2001US-271424P.
PR 15-JUN-2001; 2001US-298087P.
XX
XX
PA (UYRP ) UNIV ROCHESTER.
XX
XX
PI Zauderer M, Smith ES;
XX
XX
DR WPI; 2003-066785/06.
XX
XX
PT Selecting polynucleotides which encode antigen-specific immunoglobulin
PT molecules, by introducing the library of polynucleotides into the host
PT cells, and recovering the polynucleotides of the library for the
PT antigen -
XX
XX
PS Example 5; Page 45; 108pp; English.
XX
XX
CC The invention relates to selecting polynucleotides which encode antigen
CC -specific immunoglobulins (Ig) (or fragments) comprising introducing into
CC a population of host cells, a 1st and 2nd library of polynucleotides
CC encoding, several 1st and 2nd Ig subunit polypeptides, permitting
CC expression of Ig molecules (via control element e.g. an early/late
CC promoter), contacting Ig molecules with an antigen,
CC and recovering polynucleotides of the 1st library for the antigen.
CC The Ig molecules are heavy and light chain constant regions and
CC variable regions linked via peptide linkers and optionally directed via
CC signal peptides or transmembrane domains to different cell compartments.
CC Also included is a method of selecting polynucleotides which encode a
CC single-domain antigen-specific Ig molecule (its anti-specific fragment),
CC by: (a) introducing into a population of eukaryotic host cells capable of
CC expressing the Ig molecule a library of polynucleotides encoding
CC (through operable association with a transcriptional control region)
CC several single-domain Ig polypeptides (each comprising a Ig heavy chain
CC constant region, a camelised Ig heavy chain variable region, and a
CC signal peptide capable of directing cell surface expression or
CC secretion of Ig subunit polypeptide); (b) permitting expression of Ig
CC molecules (or antigen-specific fragments) from the host cells;
CC (c) contacting the Ig molecules with an antigen; and (d) recovering
CC polynucleotides of the library from those host cells expressing Ig
CC molecules which bind the antigens. The methods are useful for selecting
CC polynucleotides which encode an antigen-specific Ig molecule, or its
CC fragment. The present sequence is a major histocompatibility class II
CC (MHC II), cytotoxic T-lymphocyte (CTL) epitope expressed on the surface
CC of host cells used in the method of the invention.
XX
XX
SQ Sequence 8 AA;
XX
Query Match 100.0%; Score 38; DB 24; Length 8;

```

```

Beat Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   |||||
DB 1 SIINFEKL 8

RESULT 74
AAB84323
ID AAB84323 standard; peptide; 9 AA.
XX
XX
AC AAB84323;
XX
XX
DT 22-AUG-2001 (first entry)
XX
DE Peptide used to produce IemA peptide variants.
XX
XX
KW IemA; CD8+ epitope; T cell response.
XX
XX
OS Synthetic.
XX
XX
PN WO200140275-A2.
XX
XX
PD 07-JUN-2001.
XX
XX
PR 06-DEC-2000; 2000WO-US33027.
XX
XX
PR 06-DEC-1999; 99US-0169227.
XX
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX
PI Kurlander RJ, Chao E, Fields J;
XX
XX
DR WPI; 2001-389952/41.
XX
XX
PT New isolated variant of IemA, IemA, comprising a hydrophobic element
PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT response or as a treatment or prophylactic against diseases -
XX
XX
PS Disclosure; Page 7; 65pp; English.
XX
XX
CC The specification describes a peptide variant of IemA, comprising a
CC hydrophobic element joined to a CD8+ epitope. The peptides may be
CC used therapeutically by administering the peptides to a patient having
CC a need to induce a directed CD8+ T cell response. The peptide may also
CC be used as a preventive measure to avoid a disease or condition, or to
CC treat subjects already afflicted with a disease. The present sequence
CC was used to create peptides of the invention.
XX
XX
SQ Sequence 9 AA;
XX
Query Match 100.0%; Score 38; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   |||||
DB 2 SIINFEKL 9

RESULT 75
ABP57402
ID ABP57402 standard; peptide; 9 AA.
XX
XX
AC ABP57402;
XX
XX
DT 23-APR-2003 (first entry)
XX
XX
DE Synthetic 9mer peptide.
XX
XX
KW Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;

```

KM	Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW	vincidic; cytostatic; vaccine; viral infection; cancer; EtXB, CtxB.
XX	Synthetic.
OS	WO200300899-A1.
PN	03-JAN-2003.
PD	20-JUN-2002; 2002MO-CG02829.
PP	22-JUN-2001; 2001GB-0015382.
PR	(UYBR-) UNIV BRISTOL.
PA	Hirst TR;
B1	WIPI; 2003-175291/17.
DR	Use of a mutant form of B subunit of Escherichia coli heat labile enterotoxin or B subunit of cholera toxin for delivering an agent to a target cell for treating viral infection or cancer
PT	-
XX	Example 5, Page 45, 84pp; English.
PS	The present invention describes a mutant form of B subunit of Escherichia coli heat labile enterotoxin (EtXB) or B subunit of cholera toxin (CtxB) from Vibrio cholerae which is useful for delivering an agent to a target cell, and has GM-1 ganglioside receptor binding activity but has reduced immunogenic and immunomodulatory activity relative to the wild-type form of EtXB or CtxB. Also described: (1) treating a disease or condition in a subject; (2) delivering the agent using the mutant to a target cell; (3) a composition; and (4) a kit for delivering the agent to a target cell. Mutant EtXB and CtxB have virucidal and cytoprotective activities and can be used in vaccines. The mutant can be used for the preparation of a medicament for delivering an exogenous peptide, which is the agent, into the major histocompatibility complex (MHC) Class I antigen processing and presenting pathways to elicit a cytotoxic T lymphocyte (CTL) response, or for separate, simultaneous or combined use for treating viral infection or cancer. The mutant form of EtXB or CtxB enters mammalian cells without inducing a potent anti-B-subunit response and immunomodulatory response. It may be linked with an agent to upregulate the presentation of the antigen or antigenic determinant. The present sequence represents a peptide which is used in an example from the present invention.
CC	
CC	
Cc	
CO	
SO	Sequence 9 AA;
Qy	Query Match 100.0%; Score 38; DB 24; Length 9; Best Local Similarity 100.0%; Pred. No. 9.3e+05; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0 1 SIINFEKL 8 Db 2 SIINFELK 9
ID	AAM04643 standard; peptide; 10 AA. AAM04643
AC	AAM04643;
DT	01-AUG-1997 (first entry)
DE	Ovalbumin-derived activated CD8+ T cells epitope OVA10N.
KX	Macrophage; artificial antigen presenting cell; APC; cancer;
KM	tumour; neoplasia; viral infection; retroviral infection;
KW	autoimmune.
OS	Synthetic.
XX	

```

PN      MO9637107-A1.
PD      28-NOV-1996.
XX
XX      22-MAY-1996;    96WO-USO7436.
PF
PR      23-MAY-1995;    95US-0447761.
XX
XX      (SCRI ) SCRIPPS RES INST.
PA
XX      DeBrujn MLH, Jackson MR, Peterson PA;
PI
XX      WPI; 1997-020850/02.
DR
XX
XX      Prodn. of activated CD8+ T cells directed to specific antigen - can
PT specifically kill target cells useful to treat, e.g. cancer
XX
XX      Example 1, Page 26; 84pp; English.
PS
CC      The method for the production of activated CD8+ T cells specifically
CC directed towards a particular antigen involves affixing peptides
CC corresponding to the particular antigen to an artificial support;
CC contacting macrophages with the affixed peptides for a time sufficient
CC for the peptides to be engulfed, and at least a portion of the peptides
CC to be presented on the surface of the macrophage; and contacting
CC unprimed CD8+ T cells with the peptide presenting macrophages for a
CC time sufficient to activate the unprimed CD8+ T cells. The present
CC sequence represents a peptide designated OVA10N which corresponds to
CC ovalbumin, a Kb-restricted peptide antigen. This represents the optimal
CC peptide with the addition of two amino acids at th amino-terminus.
CC Small extensions to the optimal peptide affect the affinity of the
CC peptide for soluble class I molecules in vitro e.g. the addition
CC of two amino acids to the amino-terminus lowers the affinity to Kb by
CC 76-fold compared to the optimal peptide; addition of two amino acids to
CC the carboxy-terminus lowers the affinity by 4-fold. The method,
CC macrophages and artificial antigen presenting cell, having a peptide
CC corresponding to the particular antigen present on its surface and at
CC least a portion of an artificial support in its interior, can be used to
CC treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
CC infection or autoimmune or autoimmune-type conditions) in patients via
CC the specific killing of target cells.
CX
XX      Sequence 10 AA;
SQ
Query Match          100.0%; Score 38; DB 18; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 SIINFEKL 8
        |||||
Db       3 SIINFELK 10
RESULT 77
AAW04644 AAW04644 standard; peptide; 10 AA.
XX
XX      AAW04644;
AC
XX
XX      01-AUG-1997 (first entry)
DT
XX
XX      Ovalbumin-derived activated CD8+ T cells epitope OVA10C.
DE
XX      Macrophage; artificial antigen presenting cell; APC; cancer;
KW tumour; neoplasia; viral infection; retroviral infection;
KW autoimmune.
XX
XX      Synthetic.
OS
XX      WO9637107-A1.
PN
XX      28-NOV-1996.
PD

```

PF 22-MAY-1996; 96WO-US07436.
 XX
 PR 23-MAY-1995; 95US-0447761.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI DeBrujn MLH, Jackson MR, Peterson PA;
 XX
 DR WPI; 1997-020850/02.
 XX
 PT Prodn. of activated CD8+ T cells directed to specific antigen - can
 PT specifically kill target cells useful to treat, e.g. cancer
 XX
 PS Example 1; Page 26; 84pp; English.
 XX
 CC The method for the production of activated CD8+ T cells specifically
 CC directed towards a particular antigen involves affixing peptides
 CC corresponding to the particular antigen to an artificial support;
 CC contacting macrophages with the affixed peptides for a time sufficient
 CC for the peptides to be engulfed, and at least a portion of the peptides
 CC to be presented on the surface of the macrophage; and contacting
 CC unprimed CD8+ T cells with the peptide presenting macrophages for a
 CC time sufficient to activate the unprimed CD8+ T cells. The present
 CC sequence represents a peptide designated OVA10C which corresponds to
 CC ovalbumin, a Kb-restricted peptide antigen. This represents the optimal
 CC peptide with the addition of two amino acids at the carboxy-terminus.
 CC Small extensions to the optimal peptide affect the affinity of the
 CC peptide for soluble class I molecules in vitro e.g. the addition
 CC of two amino acids to the amino-terminus lowers the affinity to Kb by
 CC 76-fold compared to the optimal peptide; addition of two amino acids to
 CC the carboxy-terminus lowers the affinity by 4-fold. The method,
 CC macrophages and artificial antigen presenting cell, having a peptide
 CC corresponding to the particular antigen present on its surface and at
 CC least a portion of an artificial support in its interior, can be used to
 CC treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
 CC infection or autoimmune or autoimmune-type conditions) in patients via
 CC the specific killing of target cells.
 XX
 SQ Sequence 10 AA;
 XX
 Query Match 100.0%; Score 38; DB 18; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEXL 8
 DB 1 SIINFEXL 8
 XX
 RESULT 78
 AAU09821
 ID AAU09821 standard; peptide; 10 AA.
 XX
 AC AAU09821;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Modified ovalbumin-derived class I H-2Kb restricted peptide #1.
 XX
 KW Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX
 OS Synthetic.
 OS
 PN WO200178767-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 17-APR-2001; 2001WO-EP04313.
 XX
 PR 14-APR-2000; 2000AT-0000657.
 XX

PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Mattner F, Zauner W, Schmidt W, Buschle M;
 XX
 DR WPI; 2002-025970/03.
 XX
 PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 XX
 PS Example 1; Page 9; 18pp; English.
 XX
 CC The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEXL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDDD) at the N-terminus of peptide SIINFEXL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEXL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #1 as described in the method of the invention.
 XX
 SQ Sequence 10 AA;
 XX
 Query Match 100.0%; Score 38; DB 23; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEXL 8
 DB 3 SIINFEXL 10
 XX
 RESULT 79
 AAU09825
 ID AAU09825 standard; peptide; 10 AA.
 XX
 AC AAU09825;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Modified ovalbumin-derived class I H-2Kb restricted peptide #5.
 XX
 KW Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX
 OS Synthetic.
 OS
 PN WO200178767-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 17-APR-2001; 2001WO-EP04313.
 XX
 PR 14-APR-2000; 2000AT-0000657.
 XX
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Mattner F, Zauner W, Schmidt W, Buschle M;
 XX
 DR WPI; 2002-025970/03.
 XX
 PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 XX

PS Example 2; Page 9; 18pp; English.
 CC The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDBD) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #5 as described in the method of the invention.
 SQ Sequence 10 AA;
 OY 1 SIINFEKL 8
 DB 3 SIINFEKL 10
 RESULT 80
 AAW14122
 ID AAW14122 standard; peptide; 12 AA.
 AC AAW14122;
 DT 20-OCT-1997 (first entry)
 DE OVA protein derived MHC class I binding peptide.
 XX Major histocompatibility complex; MHC; target; binding; tumour;
 KW cancer; neoplasia; LSTRA; EL-4; identification; detection; screening;
 KW tissue typing; Bcr-abl; IFV; influenza.
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 12
 FT Misc-difference /note= "bioinylated"
 FT
 FN WO9641188-A1.
 PD 19-DEC-1996.
 PF 07-JUN-1996; 96WO-US09680.
 PR 07-JUN-1995; 95US-0485610.
 PA (UNITW) UNITV WASHINGTON.
 XX
 XX Cheever MA, Chen W;
 PI
 DR WPI; 1997-108657/10.
 PT Identifying major histocompatibility complex class I binding mols. -
 PT using peptide(s) having a core of 7-14 amino acids with extra amino
 PT acids and a reporter gp. at the N- or C-terminus, useful for tissue
 PT typing
 XX
 PS Example 3; Page 23; 41pp; English.
 CC AAW14122 is a bioinylated peptides derived from the OVA (ovalbumin)

CC protein which can be obtained from either an LSTRA or EL-4 tumour
 CC of Balb/c mice. The peptides bind to MHC class I molecules. This
 CC is useful for tissue typing or for screening for molecules that
 CC interact with MHC class I molecules. MHC class I molecules can be
 CC identified using the peptides and also the peptides are useful in
 CC vaccines against disease and infection e.g. caused by viruses,
 CC bacteria or tumours.
 SQ Sequence 12 AA;
 OY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 RESULT 81
 AAU09822
 ID AAU09822 standard; peptide; 12 AA.
 AC AAU09822;
 DT 14-FEB-2002 (first entry)
 DE Modified ovalbumin-derived class I H-2Kb restricted peptide #2.
 XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX
 OS Synthetic.
 XX
 XX WO200178767-A2.
 PD 25-OCT-2001.
 PR 17-APR-2001; 2001WO-EP04313.
 PR 14-APR-2000; 2000AT-0000657.
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 PI Matzner F, Zauner W, Schmidt W, Buschle M;
 DR WPI; 2002-025970/03.
 PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 PT
 PS Example 1; Page 9; 18pp; English.
 CC The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDBD) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #2 as described in the method of the invention.
 SQ Sequence 12 AA;

Query Match 100.0%; Score 38; DB 23; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 5 SIINFEKL 12

RESULT 82

AAU09826
 ID AAU09826 standard; peptide; 12 AA.

XX
 AC AAU09826;

DT 14-FEB-2002 (first entry)

XX Modified ovalbumin-derived class I H-2Kb restricted peptide #6.

XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;

KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;

KW popliteal lymph node; spleen; immune response; systemic response.

OS Synthetic.

PN WO200178767-A2.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-EP04313.

XX 14-APR-2000; 2000AT-0000657.

XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PI Mattner F, Zauner W, Schmidt W, Buschle M;

DR WPI; 2002-025970/03.

PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 improved immune response in a mammal, comprises a modified peptide -
 PS Example 2; Page 9; 18pp; English.

XX The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negatively charged (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EBED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #6 as described in the method of the invention.

XX Sequence 12 AA;

QY Query Match 100.0%; Score 38; DB 23; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 5 SIINFEKL 12

RESULT 83
 AAU09827
 ID AAU09827 standard; peptide; 12 AA.

XX
 AC AAU09827;

DT 14-FEB-2002 (first entry)

XX Modified ovalbumin-derived class I H-2Kb restricted peptide #7.

XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;

KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;

KW popliteal lymph node; spleen; immune response; systemic response.

OS Synthetic.

PN WO200178767-A2.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-EP04313.

XX 14-APR-2000; 2000AT-0000657.

XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PI Mattner F, Zauner W, Schmidt W, Buschle M;

DR WPI; 2002-025970/03.

PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 improved immune response in a mammal, comprises a modified peptide -
 PS Example 3; Page 10; 18pp; English.

XX The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negatively charged (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EBED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #7 as described in the method of the invention.

XX Sequence 12 AA;

QY Query Match 100.0%; Score 38; DB 23; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 84

ABB76049
 ID ABB76049 standard; Peptide; 14 AA.

XX
 AC ABB76049;

DT 12-JUL-2002 (first entry)

DE Peptide insert in CACTES-Cys-Ova.
 XX
 XX Bordetella pertussis; adenylate cyclase; CyaA; adenylcyclase;
 KM vector; drug delivery; antigen delivery; cell targeting; CD11b;
 XX CACTES-Cys-Ova.
 XX
 OS Synthetic.
 XX
 XX EP118446-A1.
 XX
 XX 20-MAR-2002.
 PD
 XX 15-SEP-2000; 2000EP-0402562.
 PF
 XX 15-SEP-2000; 2000EP-0402562.
 PR
 XX 15-SEP-2000; 2000EP-0402562.
 PA (INSP) INST PASTEUR.
 XX (CNRS) CENT NAT RECH SCT.
 PI Leclerc C, Guernonprez P, Ladant D, Guiso N, Khelaf N;
 XX
 XX WPI; 2002-354020/39.
 DR
 XX
 XX Use of Bordetella adenylcyclase to make proteinaceous vector, useful
 PT for drug or antigen delivery, selectively targets cells that express
 PT CD11b -
 XX
 PS Example A; Page 10; 34pp; English.
 XX
 XX The present sequence is a peptide that was introduced into the
 CC catalytic domain of a detoxified form of the adenylate cyclase
 CC (CyaA, or adenylcyclase) of Bordetella pertussis. A recombinant
 CC detoxified CyaA toxin, CACTES-Cys-Ova, harbouring a unique cysteine
 CC that was derived from the peptide insert, was produced. The
 CC protein was labeled on its unique cysteine, and used to detect
 CC CyaA binding to neutrophils. Experiments showed that CyaA binding
 CC to the surface of 3 myeloid cell lines of mouse or human origin, as
 CC well as to human neutrophils, was mainly mediated through the
 CC CD11b/CD18 integrin. The invention relates to the novel use of
 CC Bordetella CyaA as a proteinaceous vector for targeting a
 CC molecule of interest to the surface CD11b-expressing cells,
 CC especially dendritic cells and neutrophils. The molecule of
 CC interest is translocated in the cytosol to prime a cytotoxic T
 CC lymphocyte response. In a preferred embodiment, a peptide is
 CC inserted into in the catalytic domain of CyaA at a permissive site.
 CC The peptide may be an intracellular bacterial cell, tumour, viral,
 CC fungal or parasite cell antigen (all claimed). Alternatively, a
 CC drug, especially an antiinflammatory, is chemically coupled to
 CC CyaA for drug delivery.
 XX
 XX
 SQ Sequence 14 AA;
 QY
 Query Match 100.0%; Score 38; DB 23; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 SIINFEKL 8
 5 SIINFEKL 12
 RESULT 85
 AAU09823
 ID AAU09823 standard; peptide; 14 AA.
 XX
 XX AAU09823;
 AC
 XX
 XX 14-FEB-2002 (first entry)
 DT
 XX
 XX Modified ovalbumin-derived class I H-2Kb restricted peptide #3.
 DE
 XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KM

KM popliteal lymph node; spleen; immune response; systemic response.
 XX
 XX Synthetic.
 OS
 XX WO200178767-A2.
 XX
 XX 25-OCT-2001.
 PD
 XX 17-APR-2001; 2001WO-EP04313.
 PF
 XX 14-APR-2000; 2000AT-0000657.
 PR
 XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 PA
 XX Matner F, Zauner W, Schmidt W, Buschle M;
 PI
 XX WPI; 2002-025970/03.
 DR
 XX
 XX Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 PT
 XX
 XX Example 1; Page 9; 18pp; English.
 PS
 XX The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EBED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #3 as described in the method of the invention.
 XX
 XX
 SQ Sequence 14 AA;
 QY
 Query Match 100.0%; Score 38; DB 23; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 SIINFEKL 8
 7 SIINFEKL 14
 RESULT 86
 AAU09828
 ID AAU09828 standard; peptide; 14 AA.
 XX
 XX AAU09828;
 AC
 XX
 XX 14-FEB-2002 (first entry)
 DT
 XX
 XX Modified ovalbumin-derived class I H-2Kb restricted peptide #8.
 DE
 XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KM popliteal lymph node; spleen; immune response; systemic response.
 XX
 XX Synthetic.
 OS
 XX WO200178767-A2.
 PN
 XX 25-OCT-2001.
 PD
 XX 17-APR-2001; 2001WO-EP04313.
 PF

```

XX 14-APR-2000; 2000AT-0000657.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX PI Mattner F, Zauner W, Schmidt W, Bueschle M;
XX
XX WPI; 2002-025970/03.
XX
XX Pharmaceutical preparation for use as a potent vaccine for inducing an
XX improved immune response in a mammal, comprises a modified peptide
XX
XX Example 4; Page 11; 18pp; English.
XX
XX The invention relates to a pharmaceutical preparation comprising a
XX modified peptide, which induces an improved immune response in a mammal
XX compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
XX negatively charged (Glu), one positively charged (Lys) amino acid) was
XX rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
XX Asp, respectively. Results showed that the addition of 4 negatively-
XX charged amino acids (EDBD) at the N-terminus of peptide SIINFEKL makes
XX this peptide (in combination with poly-L-arginine) able to induce a high
XX amount of specific interferon (IFN)-gamma-producing T cells in the
XX draining (popliteal) lymph node (local response) and in the spleen
XX (systemic response). Thus, the addition of hydrophobic amino acids as
XX well as the addition of negatively charged amino acids transforms the
XX peptide SIINFEKL to a good inducer of specific T cells. The modified
XX peptides of the pharmaceutical composition induce a stronger immune
XX response in a mammal compared to wild type antigens. The present
XX sequence represents modified ovalbumin-derived class I H-2Kb restricted
XX peptide #8 as described in the method of the invention.
XX
SQ Sequence 14 AA;

Query Match 100.0%; Score 38; DB 23; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   |||||
Db 7 SIINFEKL 14

RESULT 87
AAU09824
ID AAU09824 standard; peptide; 15 AA.
XX
XX AAU09824;
XX
AC 14-FEB-2002 (first entry)
XX
DT Modified ovalbumin-derived class I H-2Kb restricted peptide #4.
XX
DE Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
XX immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
XX popliteal lymph node; spleen; immune response; systemic response.
XX
OS Synthetic.
XX
XX WO200178767-A2.
XX
XX 25-OCT-2001.
XX
XX 17-APR-2001; 2001WO-EP04313.
XX
XX 14-APR-2000; 2000AT-0000657.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX PI Mattner F, Zauner W, Schmidt W, Bueschle M;
XX
XX WPI; 2002-025970/03.
XX

```

```

PT Pharmaceutical preparation for use as a potent vaccine for inducing an
PT improved immune response in a mammal, comprises a modified peptide
XX
XX Example 1; Page 9; 18pp; English.
XX
XX The invention relates to a pharmaceutical preparation comprising a
XX modified peptide, which induces an improved immune response in a mammal
XX compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
XX negatively charged (Glu), one positively charged (Lys) amino acid) was
XX rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
XX Asp, respectively. Results showed that the addition of 4 negatively-
XX charged amino acids (EDBD) at the N-terminus of peptide SIINFEKL makes
XX this peptide (in combination with poly-L-arginine) able to induce a high
XX amount of specific interferon (IFN)-gamma-producing T cells in the
XX draining (popliteal) lymph node (local response) and in the spleen
XX (systemic response). Thus, the addition of hydrophobic amino acids as
XX well as the addition of negatively charged amino acids transforms the
XX peptide SIINFEKL to a good inducer of specific T cells. The modified
XX peptides of the pharmaceutical composition induce a stronger immune
XX response in a mammal compared to wild type antigens. The present
XX sequence represents modified ovalbumin-derived class I H-2Kb restricted
XX peptide #4 as described in the method of the invention.
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 38; DB 23; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   |||||
Db 8 SIINFEKL 15

RESULT 88
ABP57403
ID ABP57403 standard; peptide; 16 AA.
XX
XX ABP57403;
XX
DT 23-APR-2003 (first entry)
XX
DE Synthetic 16mer peptide.
XX
XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virulence; cytosolic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS Synthetic.
XX
XX WO2003000899-A1.
XX
XX 03-JAN-2003.
XX
XX 20-JUN-2002; 2002WO-GB02829.
XX
XX 22-JUN-2001; 2001GB-0015382.
XX
XX (UYBR-) UNIV BRISTOL.
XX
XX Hirst TR;
XX
XX WPI; 2003-175291/17.
XX
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer
XX
XX Example 5; Page 45; 84pp; English.
XX
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target

```


cell, and has GM-1 ganglioside receptor binding activity but has reduced immunogenic and immunomodulatory activity relative to the wild-type form of EtXB or CtxB. Also described: (1) treating a disease or condition in a subject; (2) delivering the agent using the mutant to a target cell; (3) a composition; and (4) a kit for delivering the agent to a target cell. Mutant EtXB and CtxB have virucide and cytostatic activities and can be used in vaccines. The mutant can be used for the preparation of a medicament for delivering an exogenous peptide, which is the agent, into the major histocompatibility complex (MHC) Class I antigen processing and presenting pathways to elicit a cytotoxic T lymphocyte (CTL) response, or for separate, simultaneous or combined use for treating viral infection or cancer. The mutant form of EtXB or CtxB enters mammalian cells without inducing a potent anti-B-subunit response and immunomodulatory response. It may be linked with an agent to upregulate the presentation of the antigen or antigenic determinant. The present sequence represents a peptide which is used in an example from the present invention.

XX Sequence 16 AA;

Query Match 100.0%; Score 38; DB 24; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.25; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEXL 8
|||
Db 9 SIINFEXL 16

RESULT 89

AAW19957 standard; Peptide; 19 AA.

XX AAW19957;

DT 10-NOV-1997 (first entry)

DE BiP-binding domain-OVA hybrid peptide.

KW Vaccine; immunotherapy; heat shock protein; BiP; OVA; cancer; infectious disease.

OS Synthetic.

FT Key Location/Qualifiers

FT Peptide 1..8 /label= BiP

FT Peptide 9..11 /label= Linker

FT Peptide 12..19 /label= Ova

PN WO9706821-A1.

PD 27-FEB-1997.

PF 16-AUG-1996; 96WO-US13363.

PR 18-AUG-1995; 95US-0002490.

PR 18-AUG-1995; 95US-0002479.

PA (SLOAN KETTERING INST CANCER RES.

PI Harl FU, Hoe MH, Houghton A, Mayhew M, Rothman JE;

PI Takeuchi Y;

DR WPI; 1997-165035/15.

PT Composn. for inducing immune response contg. antigen and heat shock protein - also new hybrid peptide and related nucleic acid, for treatment of infectious diseases and tumours

PS Example 1; Page 18; 58pp; English.

XX Hybrid peptides OVA-BiP (AAW19956) and BiP-OVA (AAW19957) comprise chicken OVA-peptide (see AAW19955) joined via a peptide linker to heat shock protein (HSP) BiP binding domain (see also AAW19951). The hybrid peptide can be combined in vitro with a HSP, such as hsp70, to form a complex that, when administered to a subject, induces an immune response.

XX Sequence 19 AA;

Query Match 100.0%; Score 38; DB 18; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.3; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEXL 8
|||
Db 12 SIINFEXL 19

RESULT 90

AAW19956 standard; Peptide; 19 AA.

XX AAW19956;

DT 10-NOV-1997 (first entry)

DE OVA-BiP-binding domain hybrid peptide.

KW Vaccine; immunotherapy; heat shock protein; BiP; OVA; cancer; infectious disease.

OS Synthetic.

FT Key Location/Qualifiers

FT Peptide 1..8 /label= OVA

FT Peptide 9..11 /label= Linker

FT Peptide 12..19 /label= BiP

PN WO9706821-A1.

PD 27-FEB-1997.

PF 16-AUG-1996; 96WO-US13363.

PR 18-AUG-1995; 95US-0002490.

PR 18-AUG-1995; 95US-0002479.

PA (SLOAN KETTERING INST CANCER RES.

PI Harl FU, Hoe MH, Houghton A, Mayhew M, Rothman JE;

PI Takeuchi Y;

DR WPI; 1997-165035/15.

PT Composn. for inducing immune response contg. antigen and heat shock protein - also new hybrid peptide and related nucleic acid, for treatment of infectious diseases and tumours

PS Example 1; Page 18; 58pp; English.

CC Hybrid peptides OVA-BiP (AAW19956) and BiP-OVA (AAW19957) comprise chicken OVA-peptide (see AAW19955) joined via a peptide linker to heat shock protein (HSP) BiP binding domain (see also AAW19951). The hybrid peptide can be combined in vitro with a HSP, such as hsp70, to form a complex that, when administered to a subject, induces an immune response.

XX Sequence 19 AA;

Query Match 100.0%; Score 38; DB 18; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 91

AAE13446
 ID AAE13446 standard; peptide; 19 AA.

AC AAE13446;
 XX

DT 12-FEB-2002 (first entry)
 XX

DE Chicken MHC class I peptide antigen #1.
 XX

KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.

OS Gallus gallus.
 XX

PH Key Location/Qualifiers
 FT Region 1..8 /note= "MHC class I epitope"

FT Region 9..11 /note= "Linker"

FT Region 12..19 /note= "Javelin sequence"

FT Region 12..19 /note= "Javelin sequence"

PN WO200179259-A1.
 XX

PD 25-OCT-2001.
 XX

PF 17-APR-2001; 2001WO-US12567.
 XX

PR 17-APR-2000; 2000US-197462P.
 XX

PA (ROTH/) ROTHMAN J E.
 PA (MAYH/) MAYHEW M.
 PA (HOEM/) HOE M.

PI Rothman JE, Mayhew M, Hoe M;
 XX

DR WPI; 2002-017594/02.
 XX

PT A new antigenic complex comprising epitopes non-covalently joined to a
 PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -

PS Disclosure; Page 13; 47pp; English.
 XX

CC The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken MHC class I peptide antigen.
 CC

SQ Sequence 19 AA;
 XX

Query Match 100.0%; Score 38; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||

DB 1 SIINFEXL 8

RESULT 92

AAE13447
 ID AAE13447 standard; peptide; 19 AA.

AC AAE13447;
 XX

DT 12-FEB-2002 (first entry)
 XX

DE Chicken MHC class I peptide antigen #2.
 XX

KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.

OS Gallus gallus.
 XX

PH Key Location/Qualifiers
 FT Region 1..8 /note= "Javelin sequence"

FT Region 9..11 /note= "Linker"

FT Region 12..19 /note= "MHC class I epitope"

FT Region 12..19 /note= "MHC class I epitope"

PN WO200179259-A1.
 XX

PD 25-OCT-2001.
 XX

PF 17-APR-2001; 2001WO-US12567.
 XX

PR 17-APR-2000; 2000US-197462P.
 XX

PA (ROTH/) ROTHMAN J E.
 PA (MAYH/) MAYHEW M.
 PA (HOEM/) HOE M.

PI Rothman JE, Mayhew M, Hoe M;
 XX

DR WPI; 2002-017594/02.
 XX

PT A new antigenic complex comprising epitopes non-covalently joined to a
 PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -

PS Disclosure; Page 13; 47pp; English.
 XX

CC The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken MHC class I peptide antigen.
 CC

SQ Sequence 19 AA;
 XX

Query Match 100.0%; Score 38; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 12 SIINFEXL 19

RESULT 93
 ABP57404
 ID ABP57404 standard; peptide; 19 AA.

```

XX ABP57404;
AC
XX
XX 23-APR-2003 (first entry)
DT
XX
XX Synthetic 19mer peptide.
DE
XX
XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW
KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
RM
RM virulence; cytosolic; vaccine; viral infection; cancer; EtxB; CtxB.
OS
OS Synthetic.
XX
XX WO2003000899-A1.
PN
XX
XX 03-JAN-2003.
PD
XX
XX 20-JUN-2002; 2002WO-CB02829.
PF
XX
XX 22-JUN-2001; 2001GB-0015382.
PR
XX
XX (UYBR-) UNIV BRISTOL.
PA
XX
XX Hirst TR;
PI
XX
XX WPI, 2003-175291/17.
DR
XX
XX Use of a mutant form of B subunit of Escherichia coli heat labile
PT enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT target cell for treating viral infection or cancer
XX
XX Example 5; Page 45; 84pp; English.
XX
XX The present invention describes a mutant form of B subunit of Escherichia
CC coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC from Vibrio cholerae which is useful for delivering an agent to a target
CC cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC immunogenic and immunomodulatory activity relative to the wild-type form
CC of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC a subject; (2) delivering the agent using the mutant to a target cell;
CC (3) a composition; and (4) a kit for delivering the agent to a target
CC cell. Mutant EtxB and CtxB have virulence and cytotoxic activities and
CC can be used in vaccines. The mutant can be used for the preparation of
CC a medicament for delivering an exogenous peptide, which is the agent,
CC into the major histocompatibility complex (MHC) Class I antigen
CC processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC (CTL) response, or for separate, simultaneous or combined use for
CC treating viral infection or cancer. The mutant form of EtxB or CtxB
CC enters mammalian cells without inducing a potent anti-B-subunit response
CC and immunomodulatory response. It may be linked with an agent to
CC upregulate the presentation of the antigen or antigenic determinant.
CC The present sequence represents a peptide which is used in an example
CC from the present invention.
XX
XX Sequence 19 AA;
SO
XX
XX Query Match 100.0%; Score 38; DB 24; Length 19;
XX Best Local Similarity 100.0%; Pred. No. 0.3;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX 1 SIINFELK 8
OY |||||
DB 12 SIINFELK 19
XX
XX
XX RESULT 94
XX AAR32294
ID AAR32294 standard; protein; 24 AA.
XX
XX AAR32294;
AC
XX
XX 25-MAR-2003 (updated)
DT 31-MAY-1993 (first entry)

```

[illegible]

```

XX 19-FEB-1992; 92US-0841662.
XX (SCRI ) SCRIPPS RES INST.
XX Jackson M, Langlade-demoyen P, Peterson PA;
XX WPI; 1993-288401/36.
XX Prodn. and use of human class I MHC molecules for activation of
XX CD8 cells - for therapy of e.g. cancer, viral, retroviral and
XX auto-immune diseases
XX
XX Disclosure; Page 77; 182pp; English.
XX
XX Human class I MHC genes are inserted into a cell and placed under
XX the control of an inducible promoter. This provides a means of
XX producing, loading and using Class I MHC molecules to specifically
XX activate CD8 cells in vitro. Activated cells can be used to
XX specifically kill target cells and also to treat cancer as well as
XX viral, retroviral, autoimmune and autoimmuno-type diseases. When
XX conjugated to a toxin, empty human MHC molecules expressed by the
XX cells can be used to inhibit transplant rejection. A number of
XX antigenic peptides (AAR41450-R41463) are synthesised to be bound by
XX the MHC molecules and this binding can then activate the CD8 cells.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 24 AA;

```

```

Query Match 100.0%; Score 38; DB 14; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 SIINFEKL 8
Db 5 SIINFEKL 12

```

RESULT 96

```

AAW04645
ID AAW04645 standard; peptide; 24 AA.

```

```

XX AAW04645;
XX
XX

```

```

DT 01-AUG-1997 (first entry)
XX

```

```

DE Ovalbumin-derived activated CD8+ T cells epitope OVA24.
XX

```

```

XX Macrophage; artificial antigen presenting cell; APC; cancer;
XX tumours; neoplasia; viral infection; retroviral infection;
XX autoimmuno.
XX

```

```

OS Synthetic.
XX

```

```

XX WO9637107-A1.
XX

```

```

XX 28-NOV-1996.
XX

```

```

XX 22-MAY-1996; 96WO-US07436.
XX

```

```

XX 23-MAY-1995; 95US-0447761.
XX

```

```

XX (SCRI ) SCRIPPS RES INST.
XX

```

```

XX DeBruin MLH, Jackson MR, Peterson PA;
XX

```

```

XX WPI; 1997-020850/02.
XX

```

```

XX Prodn. of activated CD8+ T cells directed to specific antigen - can
XX specifically kill target cells useful to treat, e.g. cancer
XX
XX

```

```

XX Example 1; Page 26; 84pp; English.
XX

```

```

CC The method for the production of activated CD8+ T cells specifically
CC directed towards a particular antigen involves affixing peptides
CC corresponding to the particular antigen to an artificial support;
CC contacting macrophages with the affixed peptides for a time sufficient
CC for the peptides to be engulfed, and at least a portion of the peptides
CC to be presented on the surface of the macrophage; and contacting
CC unprimed CD8+ T cells with the peptide presenting macrophages for a
CC time sufficient to activate the unprimed CD8+ T cells. The present
CC sequence represents a peptide designated OVA24 which corresponds to
CC ovalbumin, a Kb-restricted peptide antigen. This is not as efficient as
CC the optimal peptide. The method, macrophages and artificial antigen
CC presenting cell, having a peptide corresponding to the particular
CC antigen present on its surface and at least a portion of an artificial
CC support in its interior, can be used to treat conditions (e.g. cancer,
CC tumours, neoplasia, viral or retroviral infection or autoimmuno or
CC autoimmuno-type conditions) in patients via the specific killing of
CC target cells.

```

```

SQ Sequence 24 AA;

```

```

Query Match 100.0%; Score 38; DB 18; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 SIINFEKL 8
Db 5 SIINFEKL 12

```

RESULT 97

```

AAG65170
ID AAG65170 standard; peptide; 24 AA.

```

```

XX AAG65170;
XX
XX

```

```

DT 15-OCT-2002 (first entry)
XX

```

```

DE Ovalbumin based peptide.
XX

```

```

XX Cytotoxic T-lymphocyte response; CTL; antigen; viral infection;
XX bacterial infection; cancer; parasitic infection; immune response;
XX non-toxic.
XX

```

```

OS Unidentified.
XX

```

```

XX US6270769-B1.
XX

```

```

XX 07-AUG-2001.
XX

```

```

XX 24-MAY-1995; 95US-0449728.
XX

```

```

XX 24-JUL-1993; 92US-0919787.
XX

```

```

XX 25-JUL-1991; 91US-0735069.
XX

```

```

XX (IDEC-) IDEC PHARM CORP.
XX

```

```

XX Raychaudhuri S, Rastetter WH;
XX

```

```

XX WPI; 2001-564234/63.
XX

```

```

XX Induction of cytotoxic T-lymphocyte responses -
XX

```

```

XX Disclosure; Column 17-18; 24pp; English.
XX

```

```

XX The present invention relates to a method of treating viral, parasitic
XX and bacterial infections and cancer in humans, by administering an
XX antigen which causes a cytotoxic T-lymphocyte response. Said antigen does
XX not contain an immunostimulatory element. The method can also be used in
XX domesticated animals. The present sequence is a peptide used as an
XX antigen in the exemplification of the invention.
XX

```

```

XX Sequence 24 AA;
XX

```

Query Match 100.0%; Score 38; DB 22; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 5 SIINFEKL 12

RESULT 98

AA874439
ID AA874439 standard; peptide; 24 AA.

AA874439;

29-MAY-2001 (first entry)

Myelin basic protein amino acids 84-102 peptide.

Cytotoxic T-lymphocyte; CTL response; immunostimulation; infection;
cancer.

Unidentified.

US6197311-B1.

06-MAR-2001.

17-FEB-1998; 98US-0024220.

07-JUN-1995; 95US-0476674.

25-JUL-1991; 91US-0735069.

24-JUL-1992; 92US-0919787.

07-DEC-1994; 94US-0351001.

(IDEC-) IDEC PHARM CORP.

Raychaudhuri S, Raeteter WH, Black A;

WPI; 2001-256350/26.

Treating papillomavirus-related tumor or malignancy, involves
administering an antigen formulation substantially free of
immunostimulatory peptides, and comprising human papillomavirus antigen
and microfluidized adjuvant -

Disclosure; Column 10-11; 22pp; English.

The present invention describes a method of treating a
papillomavirus-related tumor, involving administering an antigen
formulation capable of inducing a cytotoxic T-lymphocyte (CTL) response
specific to the papillomavirus antigen in the individual. This is useful
in the treatment of cancer and infections, such as those due to HIV,
bacteria, parasites, influenza, herpes virus and hepatitis viruses. The
present sequence is a peptide used to demonstrate the method of the
invention.

Sequence 24 AA;

Query Match 100.0%; Score 38; DB 22; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 5 SIINFEKL 12

RESULT 99

ABG31664
ID ABG31664 standard; peptide; 24 AA.

ABG31664;

05-NOV-2002 (first entry)
Ovalbumin (OVA) peptide fragment.

Ovalbumin; microfluidised antigen; detergent; micelle-forming agent;
biodegradable oil; biocompatible oil; cytotoxic T lymphocyte; HIV;
human immunodeficiency virus; herpes virus; malaria; influenza; cancer;
hepatitis; respiratory syncytial virus; domesticated animal; OVA;
agricultural animal.

Unidentified.

US2002039582-A1.

04-APR-2002.

20-DEC-2000; 2000US-0740003.

17-FEB-1998; 98US-0024220.

25-JUL-1991; 91US-0735069.

07-DEC-1994; 94US-0351001.

07-JUN-1995; 95US-0476674.

29-AUG-1997; 97US-0919787.

(IDEC-) IDEC PHARM CORP.

Raychaudhuri S, Raeteter WH, Black A;

WPI; 2002-607062/65.

Composition useful for inducing cytotoxic T lymphocyte response in
domesticated animals and humans comprises antigen mixed with
microfluidised antigen formulation which is substantially free of
immunostimulating peptides -

Disclosure; Page 6; 31pp; English.

The invention relates to a composition comprising an antigen mixed with
microfluidised antigen formulation comprising a stabilised detergent, a
micelle-forming agent and a biodegradable/biocompatible oil. The
composition is formulated as a stable oil-in-water emulsion substantially
free of or lacking immunostimulating peptides, and is capable of inducing
specific cytotoxic T lymphocyte response against antigens in vivo. The
composition is useful for treating patients infected with human
immunodeficiency virus (HIV) or herpes virus, and patients suffering from
malaria, influenza, hepatitis, cancer or respiratory syncytial virus by
administering a composition comprising HIV antigen, malaria-associated
antigen, hepatitis-associated antigen, cancer-associated antigen, herpes
antigen or respiratory syncytial antigen respectively, mixed with
microfluidised antigen consisting essentially of two of stabilising
detergent, micelle-forming agent and biodegradable and biocompatible oil,
the antigen formulation being formulated as stable oil-in-water emulsion,
and inducing cytotoxic T lymphocyte response in a patient e.g. human,
domesticated animal or agricultural animal. This sequence represents an
ovalbumin (OVA) peptide fragment used in the scope of the invention.

Sequence 24 AA;

Query Match 100.0%; Score 38; DB 23; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 5 SIINFEKL 12

RESULT 100

ABP57405
ID ABP57405 standard; peptide; 26 AA.

ABP57405;

```

XX 23-APR-2003 (first entry)
XX
XX Synthetic 26mer peptide.
XX
XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virucide; cytosstatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
XX Synthetic.
XX
XX WO2003000899-A1.
XX
XX 03-JAN-2003.
XX
XX 20-JUN-2002; 2002WO-GB02829.
XX
XX 22-JUN-2001; 2001GB-0015382.
XX
XX (UYBR-) UNIV BRISTOL.
XX
XX Hirt TR;
XX
XX WPI; 2003-175291/17.
XX
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer
XX
XX Example 5; Page 45; 84pp; English.
XX
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytosstatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX (CTL) response, or for separate, simultaneous or combined use for
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX
XX Sequence 26 AA;
XX
XX Query Match 100.0%; Score 38; DB 24; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.41;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEXT 8
XX |||||
XX DB 19 SIINFEXT 26
XX
XX RESULT 101
XX ABP57406 standard; peptide; 26 AA.
XX
XX AC ABP57406;
XX
XX DT 23-APR-2003 (first entry)
XX
XX DE Synthetic 26mer* peptide.
XX
XX

```

```

XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virucide; cytosstatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
XX Synthetic.
XX
XX WO2003000899-A1.
XX
XX 03-JAN-2003.
XX
XX 20-JUN-2002; 2002WO-GB02829.
XX
XX 22-JUN-2001; 2001GB-0015382.
XX
XX (UYBR-) UNIV BRISTOL.
XX
XX Hirt TR;
XX
XX WPI; 2003-175291/17.
XX
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer
XX
XX Example 5; Page 45; 84pp; English.
XX
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytosstatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX (CTL) response, or for separate, simultaneous or combined use for
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX
XX Sequence 26 AA;
XX
XX Query Match 100.0%; Score 38; DB 24; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.41;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEXT 8
XX |||||
XX DB 19 SIINFEXT 26
XX
XX RESULT 102
XX AAE13448
XX ID AAE13448 standard; peptide; 30 AA.
XX
XX AC AAE13448;
XX
XX DT 12-FEB-2002 (first entry)
XX
XX DE Chicken MHC class I peptide antigen #3.
XX
XX AC Antigenic complex; epitope; heat shock protein; HSP; tether; jayelin;
XX major histocompatibility complex; MHC; therapy; immune response;
XX malignancy; chicken.
XX
XX OS Gallus gallus.
XX

```

```

XX Key Location/Qualifiers
FH Region 1..8
FT Region /note= "Javelin sequence"
FT Region 9..11 /note= "Linker"
FT Region 12..19 /note= "MHC class I epitope"
FT Region 20..22 /note= "Linker"
FT Region 23..30 /note= "Javelin sequence"
FT Region
XX WO200179259-A1.
XX 25-OCT-2001.
XX 17-APR-2001; 2001WO-US12567.
XX 17-APR-2000; 2000US-197462P.
XX (ROTH/) ROTHMAN J E.
XX (MAYH/) MAYHEW M.
XX (HOEW/) HOE M.
XX Rothman JE, Mayhew M, Hoe M;
XX WPI; 2002-017594/02.
XX A new antigenic complex comprising epitopes non-covalently joined to a
XX heat shock protein by a molecular tether designated a javelin are
XX useful to treat or prevent infectious disease or malignancy -
XX Disclosure; Page 13; 47pp; English.
XX The present invention relates to an antigenic complex, comprising a
XX number of epitopes non-covalently joined to a heat shock protein (HSP) by
XX a tethering molecule referred to as javelin which has affinity for the
XX HSP under physiological conditions, where the epitopes are covalently
XX joined to the tethering molecule and one epitope is major
XX histocompatibility complex class I (MHC) and the other MHC class II. The
XX antigenic complex is used to induce immune responses directed towards the
XX treatment or prevention of infectious diseases and malignancies. The
XX present sequence is chicken MHC class I peptide antigen.
XX Sequence 30 AA;
SQ
Query Match 100.0%; Score 38; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 12 SIINFEKL 19

```

RESULT 103

```

ID ABP57407 standard; peptide; 31 AA.
XX ABP57407;
AC 23-APR-2003 (first entry)
XX 23-APR-2003 (first entry)
XX Synthetic 31mer peptide.
XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virucide; cytostatic; vaccine; viral infection; cancer; CtxB; CtxB.
XX Synthetic.
OS
XX WO2003000899-A1.
PN

```

```

XX 03-JAN-2003.
XX 20-JUN-2002; 2002WO-GB02829.
XX 22-JUN-2001; 2001GB-0015382.
XX (UVR-) UNIV BRISTOL.
XX Hirst TR;
XX WPI; 2003-175291/17.
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer -
XX Example 5; Page 45; 84pp; English.
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX (CTL) response, or for separate, simultaneous or combined use for
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX Sequence 31 AA;
SQ
Query Match 100.0%; Score 38; DB 24; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 19 SIINFEKL 26

```

RESULT 104

```

ID AAM04646 standard; peptide; 35 AA.
XX AAM04646;
AC 01-AUG-1997 (first entry)
XX 01-AUG-1997 (first entry)
XX Ovalbumin-derived activated CD8+ T cells epitope OVA35.
XX Macrophage; artificial antigen presenting cell; APC; cancer;
XX tumours; neoplasia; viral infection; retroviral infection;
XX autoimmune.
XX Synthetic.
OS
XX WO9637107-A1.
XX 28-NOV-1996.
XX 22-MAY-1996; 96WO-US07436.
XX

```

```

PR 23-MAY-1995; 95US-0447761.
XX
XX (SCRI ) SCRIPPS RES INST.
XX
XX DeBrujn MLH, Jackson MR, Peterson PA;
XX
XX WPI, 1997-020850/02.
XX
XX Prodn. of activated CD8+ T cells directed to specific antigen - can
XX specifically kill target cells useful to treat, e.g. cancer
XX
XX Example 1; Page 26; 84pp; English.
XX
XX The method for the production of activated CD8+ T cells specifically
XX directed towards a particular antigen involves affixing peptides
XX corresponding to the particular antigen to an artificial support;
XX contacting macrophages with the affixed peptides for a time sufficient
XX for the peptides to be engulfed, and at least a portion of the peptides
XX to be presented on the surface of the macrophage; and contacting
XX unprimed CD8+ T cells with the peptide presenting macrophages for a
XX time sufficient to activate the unprimed CD8+ T cells. The present
XX sequence represents a peptide designated OVA35 which corresponds to
XX ovalbumin, a Kb-restricted peptide antigen. This is not as efficient as
XX the optimal peptide. The method, macrophages and artificial antigen
XX presenting cell, having a peptide corresponding to the particular
XX antigen present on its surface and at least a portion of an artificial
XX support in its interior, can be used to treat conditions (e.g. cancer,
XX tumours, neoplasia, viral or retroviral infection or autoimmune or
XX autoimmune-type conditions) in patients via the specific killing of
XX target cells.
XX
XX Sequence 35 AA;
SQ
Query Match 100.0%; Score 38; DB 18; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 16 SIINFEKL 23
RESULT 105
AAO26741
ID AAO26741 standard; peptide; 36 AA.
AC AAO26741;
XX
XX 27-MAR-2003 (first entry)
XX
XX Chicken ovalbumin class I and II carrying antigenic peptide.
XX
XX Cytostatic; universal polypeptidic carrier; Gb3 receptor; cytotoxic drug;
XX tumour cell; immunogenic; gene therapy; chicken.
XX
XX Gallus sp.
XX
XX EPI229045-A1.
XX
XX 07-AUG-2002.
XX
XX 01-FEB-2001; 2001EP-0400255.
XX
XX 01-FEB-2001; 2001EP-0400255.
XX
XX (CUR1-) INST CURIE.
XX (CNRS ) CENT NAT RECH SCI.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (UYPA-) UNIV CURIE PARIS VI P & M.
XX
XX Johannes L, Tartour E, Goud B, Fridman WH;
XX
XX WPI, 2003-185783/19.

```

```

XX
XX Novel universal polypeptidic carrier for targeting a molecule directly
XX or indirectly to Gb3 receptor expressing cells; useful for directing
XX cytotoxic drugs to tumor cells expressing Gb3 receptor -
XX
XX Example 3; Page 8; 24pp; English.
XX
XX The invention relates to a universal polypeptidic carrier for targeting
XX directly or indirectly a molecule to Gb3 receptor expressing cells. The
XX universal polypeptidic carrier is useful for targeting directly or
XX indirectly, to Gb3 receptor expressing cells, using molecules such as
XX proteins, peptides, oligopeptides, glycoproteins, glycopeptides, nucleic
XX acids, polynucleotides, or its combination, an antigen to be targeted
XX to antigen presenting cells or a cytotoxic drug or pro-drug to be targeted
XX to tumor cells expressing Gb3 receptor. The universal polypeptidic
XX carrier is useful for delivering an expression vector containing a
XX sequence of interest into a Gb3 receptor expressing cells which involves
XX operably linking the expression vector to a lysine-rich peptide
XX covalently linked to the Cys moiety of the universal polypeptidic
XX carrier. The lysine rich peptide is preferably a 16-mer poly-lysine and
XX the sequence of interest is preferably a sequence encoding an immunogenic
XX peptide, or a sequence encoding a drug or a pro-drug becoming toxic for
XX the Gb3 receptor expressing cells, or a sequence encoding a therapeutic
XX active molecule. The universal polypeptidic carrier is useful for
XX targeting a molecule to a Gb3 receptor expressing cell to enable the
XX molecule to be internalised, processed and/or expressed in the cell
XX as a carrier for introducing a nucleotide sequence in a target cell
XX either for gene therapy or for obtaining recombinant cells expressing
XX heterologous proteins. This sequence represents a chicken ovalbumin class
XX I and II carrying antigenic peptide relating to the universal carrier
XX protein of the invention.
XX
XX Sequence 36 AA;
SQ
Query Match 100.0%; Score 38; DB 24; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 6 SIINFEKL 13
RESULT 106
AAB84325
ID AAB84325 standard; peptide; 43 AA.
AC AAB84325;
XX
XX 22-AUG-2001 (first entry)
XX
XX Amino acid sequence of a IemS variant peptide.
XX
XX IemA; CD8+ epitope; T cell response.
XX
XX Synthetic.
XX
XX WO200140275-A2.
XX
XX 07-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-US33027.
XX
XX 06-DEC-1999; 99US-0169227.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kurlander RJ, Chao E, Fields J;
XX
XX WPI, 2001-389952/41.
XX
XX New isolated variant of IemA, comprising a hydrophobic element
XX

```


PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -
 XX
 XX
 PS Disclosure; Page 17; 65pp; English.

CC The specification describes a peptide variant of IemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC represents a peptide of the invention.
 XX
 XX

SQ Sequence 43 AA;

Query Match 100.0%; Score 38; DB 22; Length 43;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 |||||
 DB 28 SIINFEKL 35

RESULT 107

AAB84321
 ID AAB84321 standard; peptide; 47 AA.

AC AAB84321;

DT 22-AUG-2001 (first entry)

DE Amino acid sequence of a IemA variant peptide.

KM IemA; CD8+ epitope; T cell response.

OS Synthetic.

PN WO200140275-A2.

PD 07-JUN-2001.

PF 06-DEC-2000; 2000WO-US33027.

PR 06-DEC-1999; 99US-0169227.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Kurlander RJ, Chao E, Fields J;

DR WPI; 2001-389952/41.

PT New isolated variant of IemA, comprising a hydrophobic element
 PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -
 XX
 XX

PS Disclosure; Page 19; 65pp; English.

CC The specification describes a peptide variant of IemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC represents a peptide of the invention.
 XX
 XX

SQ Sequence 47 AA;

Query Match 100.0%; Score 38; DB 22; Length 47;
 Best Local Similarity 100.0%; Pred. No. 0.76;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8

DB |||||
 34 SIINFEKL 41

RESULT 108

AAB84322
 ID AAB84322 standard; peptide; 48 AA.

AC AAB84322;

DT 22-AUG-2001 (first entry)

DE Amino acid sequence of a Sema variant peptide.

KM IemA; CD8+ epitope; T cell response.

OS Synthetic.

PN WO200140275-A2.

PD 07-JUN-2001.

PF 06-DEC-2000; 2000WO-US33027.

PR 06-DEC-1999; 99US-0169227.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Kurlander RJ, Chao E, Fields J;

DR WPI; 2001-389952/41.

PT New isolated variant of IemA, comprising a hydrophobic element
 PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -
 XX
 XX

PS Disclosure; Page 19; 65pp; English.

CC The specification describes a peptide variant of IemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC represents a peptide of the invention.
 XX
 XX

SQ Sequence 48 AA;

Query Match 100.0%; Score 38; DB 22; Length 48;
 Best Local Similarity 100.0%; Pred. No. 0.78;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINFEKL 8
 |||||
 DB 1 SIINFEKL 8

RESULT 109

AAB48953
 ID AAB48953 standard; Protein; 49 AA.

AC AAB48953;

DT 27-MAR-2001 (first entry)

DE Tn5-DICE ovalbumin MHC class I epitope fusion protein.

KM Transposable element; MHC epitope; major histocompatibility complex;
 KM intracellular bacterial pathogen; IxP site; Cre recombinase;
 KM insertion end; In-frame fusion; detection; antigen;
 KM disseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
 KM ovalbumin MHC class I epitope.
 XX
 XX

```

OS Synthetic.
XX
XX WO200071158-A1.
XX
XX 30-NOV-2000.
XX
XX 26-MAY-2000; 2000WO-US14687.
XX
XX 26-MAY-1999; 99US-0136210.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Heffron FL, Parker DC, Ellefson DD;
XX
XX MPI; 2001-031967/04.
XX
XX Transposable element for detecting an antigenic epitope of a pathogen,
XX comprising 5' and 3' recombining sites, nucleic acid sequences encoding
XX a selectable marker and major histocompatibility complex (MHC) epitope,
XX and an insertion end -
XX
XX Example 2; Fig 2; 63pp; English.
XX
XX The invention relates to a novel transposable element comprising DNA
XX encoding a selectable marker (e.g., antibiotic resistance) located
XX between a 5' recombining site and a 3' recombining site (e.g., loxp
XX sites); DNA encoding an MHC (major histocompatibility complex) epitope
XX either 5' of the 5' recombining site or 3' of the 3' recombining site;
XX and insertion ends comprising an inverted repeat sequence at the 5' and
XX 3' ends of the transposable element sufficient for integration of the
XX transposable element. The transposable elements of the invention are able
XX to introduce in-frame insertions throughout the chromosome of an
XX intracellular bacterial pathogen. This system "tags" the bacterial gene
XX and resulting protein, allowing the identification of proteins
XX secreted across the membranes of the eukaryotic cell infected by the
XX bacterium. In one embodiment, the transposable elements contain an
XX antibiotic resistance cassette, two minimal loxp recombination sites, an
XX MHC class I or class II epitope, and flanking insertion ends. A
XX transposase, such as the Cre recombinase protein, is expressed in trans
XX from a plasmid, or can be included in the transposable element. The Cre
XX recombinase loops out the intervening sequences containing the antibiotic
XX resistance cassette. When the transposable element inserts within a gene,
XX the resolved insertion places the MHC class I or class II epitope in
XX frame with the gene. The transposable elements of the invention are
XX useful for detecting an antigenic epitope of an intracellular bacterial
XX pathogen, such as Salmonella sp., Mycobacterium tuberculosis and Listeria
XX monocytoenes. Certain embodiments of the technology, termed
XX "disseminated insertions of class-I epitopes" (DICE-I; DICE-II for
XX class II epitopes) allow the rapid and accurate identification of
XX proteins involved in bacterial pathogenesis so that such proteins can
XX be used as vaccine and drug targets. Carrier vaccines may be generated
XX by infecting bacteria with a transposable element of the invention
XX which additionally comprises an antigen associated with a disease,
XX preferably cancer or a viral or bacterial disease, operably linked to the
XX MHC epitope DNA of the transposable element. The present sequence
XX represents ovalbumin MHC class I epitope-containing fusion protein
XX encoded by a Tns-DICE transposable element.
XX
XX Sequence 49 AA:
XX
XX Query Match 100.0%; Score 38; DB 22; Length 49;
XX Best Local Similarity 100.0%; Pred. No. 0.8;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 STINPEKL 8
XX |||||
XX Db 12 STINPEKL 19
XX
XX RESULT 110
XX AAB48954
XX ID AAB48954 standard; Protein; 57 AA.
XX

```

```

AC AAB48954;
XX
XX 27-MAR-2001 (first entry)
XX
XX DICE-I ovalbumin MHC class I epitope-containing fusion protein.
XX
XX Transposable element; MHC epitope; major histocompatibility complex;
XX intracellular bacterial pathogen; loxp site; Cre recombinase;
XX insertion end; in-frame fusion; detection; antigen;
XX disseminated insertions of class-I epitopes; DICE-I; transposon Tns;
XX ovalbumin MHC class I epitope.
XX
XX Synthetic.
XX
XX WO200071158-A1.
XX
XX 30-NOV-2000.
XX
XX 26-MAY-2000; 2000WO-US14687.
XX
XX 26-MAY-1999; 99US-0136210.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Heffron FL, Parker DC, Ellefson DD;
XX
XX MPI; 2001-031967/04.
XX
XX N-PSDB; AAC91698.
XX
XX Transposable element for detecting an antigenic epitope of a pathogen,
XX comprising 5' and 3' recombining sites, nucleic acid sequences encoding
XX a selectable marker and major histocompatibility complex (MHC) epitope,
XX and an insertion end -
XX
XX Example 2; Fig 12; 63pp; English.
XX
XX The invention relates to a novel transposable element comprising DNA
XX encoding a selectable marker (e.g., antibiotic resistance) located
XX between a 5' recombining site and a 3' recombining site (e.g., loxp
XX sites); DNA encoding an MHC (major histocompatibility complex) epitope
XX either 5' of the 5' recombining site or 3' of the 3' recombining site;
XX and insertion ends comprising an inverted repeat sequence at the 5' and
XX 3' ends of the transposable element sufficient for integration of the
XX transposable element. The transposable elements of the invention are able
XX to introduce in-frame insertions throughout the chromosome of an
XX intracellular bacterial pathogen. This system "tags" the bacterial gene
XX and resulting protein, allowing the identification of proteins
XX secreted across the membranes of the eukaryotic cell infected by the
XX bacterium. In one embodiment, the transposable elements contain an
XX antibiotic resistance cassette, two minimal loxp recombination sites, an
XX MHC class I or class II epitope, and flanking insertion ends. A
XX transposase, such as the Cre recombinase protein, is expressed in trans
XX from a plasmid, or can be included in the transposable element. The Cre
XX recombinase loops out the intervening sequences containing the antibiotic
XX resistance cassette. When the transposable element inserts within a gene,
XX the resolved insertion places the MHC class I or class II epitope in
XX frame with the gene. The transposable elements of the invention are
XX useful for detecting an antigenic epitope of an intracellular bacterial
XX pathogen, such as Salmonella sp., Mycobacterium tuberculosis and Listeria
XX monocytoenes. Certain embodiments of the technology, termed
XX "disseminated insertions of class-I epitopes" (DICE-I; DICE-II for
XX class II epitopes) allow the rapid and accurate identification of
XX proteins involved in bacterial pathogenesis so that such proteins can
XX be used as vaccine and drug targets. Carrier vaccines may be generated
XX by infecting bacteria with a transposable element of the invention
XX which additionally comprises an antigen associated with a disease,
XX preferably cancer or a viral or bacterial disease, operably linked to the
XX MHC epitope DNA of the transposable element. The present sequence
XX represents an ovalbumin MHC class I epitope-containing fusion protein
XX encoded by the resolved sequence of a DICE-I transposable element.
XX
XX Sequence 57 AA:
XX

```

Query Match 100.0%; Score 38; DB 22; Length 57;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 DB 10 SIINFEKL 17

RESULT 111

AAE13458
 ID AAE13458 standard; Protein; 100 AA.

XX AAE13458;

DT 12-FEB-2002 (first entry)

XX Chicken ovalbumin derived protein domain #1.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;

KW major histocompatibility complex; MHC; therapy; immune response;

XX malignancy; chicken.

OS Gallus gallus.

XX Key

FT Location/Qualifiers

FT 1..8

FT Region /note= "Javelin sequence"

FT 66..73

FT Region /note= "MHC class I epitope"

FT 74..89

FT Region /note= "MHC class II epitope"

FT WO200179259-A1.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-US12567.

XX 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

PA (MAYH/) MAYHEW M.

PA (HOEM/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

PI WPI; 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a

PT heat shock protein by a molecular tether designated a javelin are

PT useful to treat or prevent infectious disease or malignancy -

PS Disclosure; Page 14; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a

CC number of epitopes non-covalently joined to a heat shock protein (HSP) by

CC a tethering molecule referred to as javelin which has affinity for the

CC HSP under physiological conditions, where the epitopes are covalently

CC joined to the tethering molecule and one epitope is major

CC histocompatibility complex class I (MHC) and the other MHC class II. The

CC antigenic complex is used to induce immune responses directed towards the

CC treatment or prevention of infectious diseases and malignancies. The

CC present sequence is chicken ovalbumin derived protein domain.

CC Sequence 100 AA;

QY Query Match 100.0%; Score 38; DB 23; Length 100;

Best Local Similarity 100.0%; Pred. No. 1.7; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

DB 58 SIINFEKL 65

RESULT 113

AAE13459

ID AAE13459 standard; Protein; 103 AA.

DB 66 SIINFEKL 73

RESULT 112

AAE13460
 ID AAE13460 standard; Protein; 100 AA.

XX AAE13460;

DT 12-FEB-2002 (first entry)

XX Chicken ovalbumin derived protein domain #3.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;

KW major histocompatibility complex; MHC; therapy; immune response;

XX malignancy; chicken.

OS Gallus gallus.

XX Key

FT Location/Qualifiers

FT 58..65

FT Region /note= "MHC class I epitope"

FT 66..81

FT Region /note= "MHC class II epitope"

FT 93..100

FT Region /note= "Javelin sequence"

FT WO200179259-A1.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-US12567.

XX 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

PA (MAYH/) MAYHEW M.

PA (HOEM/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

PI WPI; 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a

PT heat shock protein by a molecular tether designated a javelin are

PT useful to treat or prevent infectious disease or malignancy -

PS Disclosure; Page 14; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a

CC number of epitopes non-covalently joined to a heat shock protein (HSP) by

CC a tethering molecule referred to as javelin which has affinity for the

CC HSP under physiological conditions, where the epitopes are covalently

CC joined to the tethering molecule and one epitope is major

CC histocompatibility complex class I (MHC) and the other MHC class II. The

CC antigenic complex is used to induce immune responses directed towards the

CC treatment or prevention of infectious diseases and malignancies. The

CC present sequence is chicken ovalbumin derived protein domain.

CC Sequence 100 AA;

QY Query Match 100.0%; Score 38; DB 23; Length 100;

Best Local Similarity 100.0%; Pred. No. 1.7; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

DB 58 SIINFEKL 65

RESULT 113

AAE13459

ID AAE13459 standard; Protein; 103 AA.

```

XX AAE13459;
AC
XX 12-FEB-2002 (first entry)
DT
XX Chicken ovalbumin derived protein domain #2.
DE
XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KM major histocompatibility complex; MHC; therapy; immune response;
KM malignancy; chicken.
XX Gallus gallus.
OS
XX Key Location/Qualifiers
FH 1..8
FT Region /note= "Javelin sequence"
FT 9..11
FT Region /note= "Linker"
FT 69..76
FT Region /note= "MHC class I epitope"
FT 77..92
FT Region /note= "MHC class II epitope"
XX WO200179259-A1.
XX 25-OCT-2001.
XX 17-APR-2001; 2001WO-US12567.
XX 17-APR-2000; 2000US-197462P.
XX (ROTH/) ROTHMAN J E.
XX (MAYH/) MAYHEW M.
XX (HOEM/) HOE M.
XX Rothman JE, Mayhew M, Hoe M;
XX WPI; 2002-017594/02.
XX A new antigenic complex comprising epitopes non-covalently joined to a
PT heat shock protein by a molecular tether designated a javelin are
PT useful to treat or prevent infectious disease or malignancy -
XX Disclosure; Page 14; 47pp; English.
XX The present invention relates to an antigenic complex, comprising a
CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC a tethering molecule referred to as javelin which has affinity for the
CC HSP under physiological conditions, where the epitopes are covalently
CC joined to the tethering molecule and one epitope is major
CC histocompatibility complex class I (MHC) and the other MHC class II. The
CC antigenic complex is used to induce immune responses directed towards the
CC treatment or prevention of infectious diseases and malignancies. The
CC present sequence is chicken ovalbumin derived protein domain.
XX SQ Sequence 103 AA;
QY
Query Match 100.0%; Score 38; DB 23; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 69 STINFEKL 76
RESULT 114
AAE13461
ID AAE13461 standard; Protein; 103 AA.
AC AAE13461;
XX
XX 12-FEB-2002 (first entry)
DT

```

```

XX Chicken ovalbumin derived protein domain #4.
DE
XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KM major histocompatibility complex; MHC; therapy; immune response;
KM malignancy; chicken.
XX Gallus gallus.
OS
XX Key Location/Qualifiers
FH 58..65
FT Region /note= "MHC class I epitope"
FT 66..81
FT Region /note= "MHC class II epitope"
FT 93..95
FT Region /note= "Linker sequence"
FT 96..103
FT Region /note= "Javelin sequence"
XX WO200179259-A1.
XX 25-OCT-2001.
XX 17-APR-2001; 2001WO-US12567.
XX 17-APR-2000; 2000US-197462P.
XX (ROTH/) ROTHMAN J E.
XX (MAYH/) MAYHEW M.
XX (HOEM/) HOE M.
XX Rothman JE, Mayhew M, Hoe M;
XX WPI; 2002-017594/02.
XX A new antigenic complex comprising epitopes non-covalently joined to a
PT heat shock protein by a molecular tether designated a javelin are
PT useful to treat or prevent infectious disease or malignancy -
XX Disclosure; Page 14; 47pp; English.
XX The present invention relates to an antigenic complex, comprising a
CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC a tethering molecule referred to as javelin which has affinity for the
CC HSP under physiological conditions, where the epitopes are covalently
CC joined to the tethering molecule and one epitope is major
CC histocompatibility complex class I (MHC) and the other MHC class II. The
CC antigenic complex is used to induce immune responses directed towards the
CC treatment or prevention of infectious diseases and malignancies. The
CC present sequence is chicken ovalbumin derived protein domain.
XX SQ Sequence 103 AA;
QY
Query Match 100.0%; Score 38; DB 23; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 58 STINFEKL 65
RESULT 115
AAR89966
ID AAR89966 standard; Protein; 106 AA.
AC AAR89966;
XX
XX 25-MAR-2003 (updated)
DT 12-SEP-1996 (first entry)
XX Polypeptide sequence.
XX

```

Accession	Source	Protein	Region	Location/Qualifiers
XX	OS	Synthetic.		
XX	FX			
FT	FT	Key		Location/Qualifiers
FT	FT	Region		4..12
FT	FT			/note= "cytotoxic T lymphocyte (CTL) epitope #1, isolated from influenza nuclear protein (residues 366-374) "
FT	FT	Region		13..20
FT	FT			/note= "CTL epitope #2, isolated from ovalbumin (residues 257-264) "
FT	FT	Region		21..29
FT	FT			/note= "CTL epitope #3, isolated from influenza nuclear protein (residues 147-155) "
FT	FT	Region		30..37
FT	FT			/note= "CTL epitope #4, isolated from influenza nuclear protein (residues 50-58) "
FT	FT	Region		38..46
FT	FT			/note= "CTL epitope #5, isolated from murine cytomegalovirus pp89 (residues 168-176) "
FT	FT	Region		48..58
FT	FT			/note= "CTL epitope #6, isolated from adenovirus 5 E1A (residues 234-243) "
FT	FT	Region		59..67
FT	FT			/note= "CTL epitope #7, isolated from sendai virus nuclear protein (residues 324-332) "
FT	FT	Region		68..76
FT	FT			/note= "CTL epitope #8, isolated from P.Berghel circumporozoite protein (residues 249-257) "
FT	FT	Region		77..85
FT	FT			/note= "CTL epitope #9, isolated from influenza NS1 (residues 152-160) "
FT	FT	Misc-difference		78
FT	FT			/note= "encoded by GAC" "
FT	FT	Region		86..94
FT	FT			/note= "CTL epitope #10, isolated from lymphocytic choriomeningitis virus nuclear protein (residues 118-126) "
FT	FT	Region		97..106
FT	FT			/note= "monoclonal antibody epitope" "
FT	FT	Misc-difference		105
FT	FT			/note= "encoded by AGA" "
XX	XX			
PN	PN	W09603144-A1.		
XX	XX			
PD	PD	08-FEB-1996.		
XX	XX			
PF	PF	27-JUL-1995;		95WO-AU00461.
XX	XX			
PR	PR	08-FEB-1995;		95AU-0001009.
PR	PR	27-JUL-1994;		95AU-0007079.
XX	XX			
PA	PA	(COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.		
PA	PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.		
PA	PA	(UYME) UNIV MELBOURNE.		
PA	PA	(HALL-) HALL INST MEDICAL RES WALTER & ELIZA.		
PA	PA	(BIOT-) BIOTECH AUSTRALIA PTY LTD.		
XX	XX	(CSLC-) CSL LTD.		
PI	PI	BUTTONS SR, Coupar BEH, Khanna R, Moss DJ, Sunbrier A, Thomson SA;		
XX	XX			
DR	DR	WPI; 1996-116788/12.		
XX	XX	N-PSDB; AAT12413.		
PT	PT	New poly:epitope cytotoxic T lymphocyte vaccines - comprising a recombinant protein including CTL epitope(s) from pathogens, free of natural flanking sequences		

XX Claim 5; Fig 5; 46pp; English.

PS

CC This sequence represents a polypeptide encoded by a DNA insert of a
CC recombinant vaccinia virus of the invention. This sequence contains 10
CC murine cytotoxic T lymphocyte (CTL) epitopes. Each of the epitopes in
CC this sequence are capable of producing a primary CTL response in mice
CC with the appropriate major histocompatibility complex (MHC) allele. This
CC sequence (and the DNA encoding it) can be used in vaccines against
CC multiple epitopes derived from several different pathogens. The vaccine
CC could alternatively contain a large number of epitopes from one pathogen
CC so that HLA diversity of the target population is covered. The vaccines
CC can be delivered by vaccinia virus, avipox virus, bacterial or
CC rhadovirus vectors, or by virus-like particles. The proteins are
CC preferably administered with ISCOMs when they are delivered directly.
CC The advantage with these vaccines is that they provide a more diverse
CC immune response.
CC (Updated on 25-MAR-2003 to correct PA field.)
CX

SQ Sequence 106 AA;

Query Match 100.0%; Score 38; DB 17; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0,

OY 1 SIINFEXL 8
Db 13 SIINFEXL 20

RESULT 116
AAE13462
ID AAE13462 standard; Protein; 108 AA.
AC AAE13462;
DT 12-FEB-2002 (first entry)
DE Chicken ovalbumin derived protein domain #5.
DX
XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW major histocompatibility complex; MHC; therapy; immune response;
KM malignancy; chicken.
OS Gallus gallus.
CS
XX
FH Key Location/Qualifiers
FT Region 1..8 /note= "Javelin sequence"
FT Region 66..73 /note= "MHC class I epitope"
FT FT /note= "MHC class I epitope"
FT Region 74..89 /note= "MHC class II epitope"
FT Region 101..108 /note= "Javelin sequence"
FT
XX WO200179259-A1.
PN
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-US12567.
XX
PR 17-APR-2000; 2000US-197462P.
XX
PA (ROTH/) ROTHMAN J E.
PA (MAYH/) MAYHEW M.
XX (HOEW/) HOE M.
XX
PI Rothman JE, Mayhew M, Hoe M;
DR WPI: 2002-017594/02.
XX
A new antigenic complex comprising epitopes non-covalently joined to a

```

PT      heat shock protein by a molecular tether designated a javelin are
PT      useful to treat or prevent infectious disease or malignancy -
XX
XX      Disclosure; Page 15; 47pp; English.
PS
XX      The present invention relates to an antigenic complex, comprising a
CC      number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC      a tethering molecule referred to as javelin which has affinity for the
CC      HSP under physiological conditions, where the epitopes are covalently
CC      joined to the tethering molecule and one epitope is major
CC      histocompatibility complex class I (MHC) and the other MHC class II. The
CC      antigenic complex is used to induce immune responses directed towards the
CC      treatment or prevention of infectious diseases and malignancies. The
CC      present sequence is chicken ovalbumin derived protein domain.
XX
XX      Sequence      108 AA;
SQ
XX
XX      Query Match      100.0%; Score 38; DB 23; Length 108;
XX      Best Local Similarity      100.0%; Pred. No. 1,8;
XX      Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0.
OY      1 SIINFEKL 8
XX      |||||
DB      66 SIINFEKL 73
XX
XX      RESULT 117
XX      AAEL13463
XX      ID      AAEL13463 standard; Protein; 111 AA.
XX      AAEL13463;
XX
XX      12-FEB-2002 (first entry)
XX
XX      Chicken ovalbumin derived protein domain #6.
DE
XX
XX      Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW      major histocompatibility complex; MHC; therapy; immune response;
XX      malignancy; chicken.
XX
XX      Gallus gallus.
OS
XX
XX      Key      Location/Qualifiers
FH      Region      1..8
FT      /note= "Javelin sequence"
FT      Region      9..11
FT      /note= "Linker"
FT      Region      69..76
FT      /note= "MHC class I epitope"
FT      Region      77..92
FT      /note= "MHC class II epitope"
FT      Region      104..111
FT      /note= "Javelin sequence"
XX
XX      WO200179259-A1.
XX
XX      25-OCT-2001.
XX
XX      17-APR-2001; 2001MO-US12567.
XX
XX      17-APR-2000; 2000US-197462P.
XX
XX      (ROTH/) ROTHMAN J E.
XX      (MAYH/) MAYHEW M.
XX      (HOEM/) HOE M.
XX
XX      Rothman JE, Mayhew M, Hoe M;
XX
XX      WPI; 2002-017594/02.
XX
XX      A new antigenic complex comprising epitopes non-covalently joined to a
XX      heat shock protein by a molecular tether designated a javelin are
XX      useful to treat or prevent infectious disease or malignancy -

```

The present invention relates to an antigenic complex, comprising a number of epitopes non-covalently joined to a heat shock protein (HSP) by a tethering molecule referred to as Javelin which has affinity for the HSP under physiological conditions, where the epitopes are covalently joined to the tethering molecule and one epitope is a major histocompatibility complex class I (MHC) and the other MHC class II. The antigenic complex is used to induce immune responses directed towards the treatment or prevention of infectious diseases and malignancies. The present sequence is chicken ovalbumin derived protein domain.

Seq Sequence 111 AA;

Query Match	100.0%	Score 38	DB 23	Length 111
Best local similarity	100.0%	Pred. No. 1.9		
Matches 8	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 1 SIINFEKL 8
 |||||
 DB 69 SIINFEKL 76

RESULT 118
 AAY52575
 ID AAY52575 standard; Protein, 132 AA.
 AC AAY52575;
 XX 28-FEB-2000 (first entry)
 DT
 DE Amino acid sequence of AOS minigene insert of expression vector pMIN.0.
 XX
 XX
 XX Chimeric; pan DR epitope; expression vector;
 KW promoter; major histocompatibility complex; MHC; targeting; peptide;
 KW epitope; antigen; presentation; class I; cytotoxic pathway;
 KW endoplasmic reticulum; class II; extracellular antigen;
 KW cytotoxic T lymphocyte; CTL; immune response; immunogenicity; assay;
 KW vaccinia; immunology; infection; pathogen; virus; HIV; HBV; HCV;
 KW hepatitis B; hepatitis C; bacterium; protozoan; tumour cell;
 KW autoimmune disease; activation; antiviral; antimalarial;
 KW immunoprotective; minigene.
 XX
 XX
 OS Synthetic.
 XX
 XX
 PN WO9958658-A2.
 PD 18-NOV-1999.
 XX
 XX
 PF 13-MAY-1999; 99WO-US10646.
 XX
 PR 13-MAY-1998; 98US-0078904.
 PR 15-MAY-1998; 98US-0085751.
 XX
 XX
 PA (EPIM-) EPIMMUNE INC.
 XX
 PI Fikes JD, Hermanson GG, Sette A, Ishioka GY, Livingston B;
 PI Chenut RW;
 DR WPI, 2000-039103/03.
 DR N-PSDB; AA238634.
 XX
 PT Expression vectors encoding major histocompatibility targeting
 PT sequence, used as, e.g. tumor vaccines -
 XX
 XX Example 1; Fig 20; 130pp; English.
 XX
 XX This sequence represents the amino acid sequence of the AOS minigene
 CC insert of the expression vector pMIN.0 (AA238634). This insert encodes
 CC several MHC class I epitopes, and also the universal MHC
 CC class II (helper T) epitope, pan DR epitope (PADRE), and was used

in an exemplification of the present invention. The invention relates to a novel expression vector comprising a promoter operably linked to a fusion gene encoding a major histocompatibility complex (MHC) targeting sequence, and two or more heterologous peptide epitopes. The MHC targeting sequence may be a class I targeting sequence, which directs an MHC class I epitope to a cytosolic pathway or to the endoplasmic reticulum, or an MHC class II targeting sequence, which directs extracellular antigens to enter the endocytic pathway to be processed into antigen peptides for presentation on MHC class II molecules. The heterologous epitopes may comprise either helper T lymphocyte (HTL) epitopes, or a cytotoxic T lymphocyte (CTL) epitope and a universal HTL epitope, such as a pan DR epitope (PADRE). The vectors are useful for stimulating an immune response in vivo, as well as for use in assaying the human immunogenicity of a human T cell peptide epitope in vivo in a non-human mammal. They provide a nucleic acid vaccine for enhancing immunity against infectious pathogens, such as viruses (e.g., HIV, hepatitis B (HBV) and hepatitis C (HCV)) bacteria, protozoa (e.g., Plasmodium falciparum, the cause of malaria) and also tumor cells and autoimmune diseases. Universal MHC class II epitopes are advantageously combined with other MHC class I and class II epitopes to increase the number of cells that are activated in response to a given antigen and provide a broader population coverage of MHC-reactive alleles.

SQ Sequence 132 AA;

Query Match 100.0%; Score 38; DB 21; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 STINFEKL 8
| | | | |
Db 75 STINFEKL 82

RESULT 119

AAE13435 AAE13435 standard; Protein; 386 AA.

AC AAE13435;

DT 12-FEB-2002 (first entry)

DE Chicken ovalbumin containing plurality of epitopes.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;

KW major histocompatibility complex; MHC; therapy; immune response;

KM malignancy; chicken.

OS Gallus gallus.

XX Key Location/Qualifiers

FT Domain 200..291

FT Region 258..265

FT Region 266..281

FT Region /note="MHC class II epitope"

PN WO200179259-A1.

XX 25-OCT-2001.

PD 17-APR-2001; 2001WO-US12567.

PF 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

PA (MAYH/) MAYHEW M.

PA (HOEM/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

XX WPI, 2002-017594/02.

DR N-PSDB; AAD22407.

XX A new antigenic complex comprising epitopes non-covalently joined to a
PT heat shock protein by a molecular tether designated a javelin are
PT useful to treat or prevent infectious disease or malignancy -
PS Example; Fig 2; 47pp; English.

The present invention relates to an antigenic complex, comprising a number of epitopes non-covalently joined to a heat shock protein (HSP) by a tethering molecule referred to as javelin which has affinity for the HSP under physiological conditions, where the epitopes are covalently joined to the tethering molecule and one epitope is major histocompatibility complex class I (MHC) and the other MHC class II. The antigenic complex is used to induce immune responses directed towards the treatment or prevention of infectious diseases and malignancies. The present sequence is chicken ovalbumin containing plurality of epitopes.

SQ Sequence 386 AA;

Query Match 100.0%; Score 38; DB 23; Length 386;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 STINFEKL 8
| | | | |
Db 258 STINFEKL 265

RESULT 120

AAAB31545 AAAB31545 standard; Protein; 409 AA.

AC AAAB31545;

DT 30-APR-2001 (first entry)

DE Amino acid sequence of chicken ovalbumin.

XX Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;

KW lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;

KM ovalbumin.

OS Gallus sp.

XX WO200104344-A2.

XX 18-JAN-2001.

XX 10-JUL-2000; 2000WO-US18828.

XX 08-JUL-1999; 99US-0143757.

XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

XX Siegel M, Chu NR, Mizzen LA;

XX WPI, 2001-138361/14.

XX N-PSDB; AAF25127.

XX Screening for compounds that stimulate Th1-like responses in CD4+ T

XX lymphocyte cells -

XX Example 8; Fig 6; 88pp; English.

The present sequence represents an ovalbumin protein. Ovalbumin was fused to a heat shock protein (HSP), and used used in the method of the invention. The specification describes a method of determining whether a compound stimulates a Th1-like response. Th1 cells are a subset of CD4+ T lymphocyte cells. The method comprises contacting naive lymphocytes in vitro with a fusion protein comprising at least a fragment of Hsp, and then detecting the Th1-like response exhibited by the cell sample. The proteins which may be used in the method of the invention are Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be

CC used to identify compounds that stimulate Th1-like responses in response
 CC to microbial pathogens.

XX Sequence 409 AA;

Query Match 100.0%; Score 38; DB 22; Length 409;

Best Local Similarity 100.0%; Pred. No. 7.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINEFKL 8
 |||||

DB 281 SIINEFKL 288

RESULT 121

ID AAE13112 standard; Protein; 479 AA.

AC AAE13112;

DT 28-JAN-2002 (first entry)

DE Human HER300-rGM-CSF fusion construct comprising OVA-derived peptide.

XX Immunostimulatory fusion protein; IFP; antigen component; therapy;
 KW immunostimulatory component; T-cell mediated immune response; DC;
 KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
 KW PAP protein; Ala Arg linker; membrane distal extracellular domain;
 KW membrane distal intracellular domain; C-terminal tag; human; GM-CSF;
 KW HER-2 protein; granulocyte-macrophage colony stimulating factor;
 KW ovalbumin-derived octapeptide; OVA; rat; HER300-rGM-CSF fusion protein.

OS Chimeric - Homo sapiens.

OS Chimeric - Rattus norvegicus.

OS Chimeric - unidentified.

PN WO200174855-A2.

PD 11-OCT-2001.

PR 30-MAR-2001; 2001WO-US10515.

PR 30-MAR-2000; 2000US-193504P.

PA (DEND-) DENDREON CORP.

PI Laus R, Vidovic D, Graddis T;

DR WPI; 2001-662365/76.

DR N-PSDB; AAD21568.

PT An immunostimulatory fusion protein comprising the intracellular domain
 PT of HER-2 and an antigen elicits an immune response to the antigen and
 PT is useful for the treatment of associated cancer associated -

PS Example 1; Page 27; 59pp; English.

CC The invention relates to immunostimulatory fusion proteins (IFP) and
 CC nucleic acid molecules encoding such proteins. The IFPs comprise a
 CC polypeptide antigen component and an immunostimulatory component derived
 CC from the intracellular domain of HER-2 protein which is effective to
 CC elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
 CC immune response to the antigen. IFP or superactivated dendritic cells
 CC are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
 CC associated with a particularly antigen. The present sequence is HER300
 CC rGM-CSF fusion protein construct which comprises human PAP
 CC signal sequence, mature PAP protein, an Ala Arg linker, human HER-2
 CC signal sequence, mature HER-2 membrane distal extracellular domain,
 CC an Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
 CC an Ala linker, a HER-2 membrane distal intracellular domain, a mature
 CC rat granulocyte-macrophage colony stimulating factor (GM-CSF) sequence
 CC and a C-terminal tag.

XX Sequence 479 AA;

Query Match 100.0%; Score 38; DB 22; Length 479;

Best Local Similarity 100.0%; Pred. No. 8.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINEFKL 8
 |||||

DB 331 SIINEFKL 338

RESULT 122

ID AAU99725 standard; Protein; 541 AA.

AC AAU99725;

DT 07-OCT-2002 (first entry)

DE Yeast/mouse SS-OVA-Kb/beta2m-c-myc-AGA2 fusion protein.

XX Mutant major histocompatibility complex class I chimeric protein; MHC;
 KW lymphocyte; T-cell receptor; tissue sample; biopsy material; pathogen;
 KW bodily fluid; T lymphocyte; neoplastic cell; tumour cell; MHC antigen;
 KW virus; protozoan; bacteria; fungi; nematode; immune response; activator;
 KW enhancer; T cell activator; mouse; recombinant yeast cell; Kb; OVA;
 KW beta2m; dev8; AGA2; SIYK; fusion protein.

OS Chimeric - Mus sp.

OS Chimeric - Saccharomyces cerevisiae.

OS Synthetic.

PN WO200246399-A2.

PD 13-JUN-2002.

PR 10-DEC-2001; 2001WO-US47817.

PR 08-DEC-2000; 2000US-254495P.

PA (UNIT) UNIV ILLINOIS FOUND.

PI Kranz DM, Brophy S;

DR WPI; 2002-527916/56.

DR N-PSDB; ABE87870.

PT New isolated mutant major histocompatibility complex class I chimeric
 PT protein displayed on surfaces of recombinant yeast cells, has improved
 PT stability, and is useful for activating immune response -

PS Example 7; Page 38-39; 96pp; English.

CC The present invention relates to a new mutant major histocompatibility
 CC complex (MHC) class I chimeric protein. The protein of the invention
 CC comprises a portion mediating binding to surfaces of recombinant yeast
 CC cells and a portion comprising peptide binding region of MHC class I
 CC protein, where the invention is improved in stability as compared with
 CC MHC class I chimeric protein which is not a mutant chimeric protein.
 CC The protein, further comprising a detectable label, is useful for
 CC detecting a lymphocyte having a T-cell receptor protein in a biological
 CC sample such as cells, tissue sample, biopsy material or bodily fluids.
 CC The method is useful for detecting a T lymphocyte that is specific for
 CC a neoplastic cell, a tumour cell, a virus-infected cell, a protozoan-
 CC infected cell, a bacterium-infected cell or a fungus-infected cell. The
 CC protein of the invention can be used to directly activate T cells, in
 CC order to identify/screen for peptide-MHC antigens. The protein is also
 CC useful in activating T cells that participate in the removal of target
 CC cells including neoplastic cells and cells infected with pathogenic
 CC agents including viruses, protozoans, bacteria, fungi or nematodes.
 CC The invention is improved in stability as compared with MHC class I
 CC protein which is not a mutant chimeric protein. The present amino acid
 CC sequence represents a chimeric MHC protein that is encoded by a

CC yeast/mouse fusion gene, as described above.
 XX Sequence 541 AA;

Query Match 100.0%; Score 38; DB 23; Length 541;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINEKL 8
 |||||
 DB 34 SIINEKL 41

RESULT 123
 AAE13110
 ID AAE13110 standard; Protein; 564 AA.

XX AAE13110;
 DT 28-JAN-2002 (first entry)

DE Human HER500 fusion protein construct comprising OVA-derived octapeptide.

XX Immunostimulatory fusion protein; IFP; antigen component; therapy;
 KW immunostimulatory component; T-cell mediated immune response; DC;
 KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
 KW PAP protein; Ala Arg linker; membrane distal extracellular domain;
 KW membrane distal intracellular domain; C-terminal tag; human; OVA;
 KW HER-2 protein; ovalbumin-derived octapeptide; HER500 fusion protein.

XX OS Chimeric - Homo sapiens.
 OS Chimeric - unidentified.

PN WO200174855-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US10515.

XX 30-MAR-2000; 2000US-193504P.

XX (DEND-) DENDREON CORP.

PI Laus R, Vidovic D, Graddis T;

XX WPI; 2001-662965/76.

DR N-PSDB; AAD21566.

PT An immunostimulatory fusion protein comprising the intracellular domain
 of HER-2 and an antigen elicits an immune response to the antigen and
 is useful for the treatment of associated cancer associated -
 Claim 7; Page 26; 59pp; English.

CC The invention relates to immunostimulatory fusion proteins (IFP) and
 CC nucleic acid molecules encoding such proteins. The IFPs comprise a
 CC polypeptide antigen component and an immunostimulatory component derived
 CC from the intracellular domain of HER-2 protein which is effective to
 CC elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
 CC immune response to the antigen. IFP or superactivated dendritic cells
 CC are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
 CC associated with a particularly antigen. The present sequence is HER500
 CC fusion protein construct which comprises human PAP signal
 CC sequence, mature PAP protein, an Ala Arg linker, human HER-2 signal
 CC sequence, mature HER-2 membrane distal extracellular domain, an
 CC Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
 CC HER-2 membrane distal intracellular domain and a C-terminal tag.

XX Sequence 564 AA;

Query Match 100.0%; Score 38; DB 22; Length 564;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINEKL 8
 |||||
 DB 331 SIINEKL 338

RESULT 124
 AAE13111
 ID AAE13111 standard; Protein; 697 AA.

XX AAE13111;
 DT 28-JAN-2002 (first entry)

DE Human HER500-rGM-CSF fusion construct comprising OVA-derived peptide.

XX Immunostimulatory fusion protein; IFP; antigen component; therapy;
 KW immunostimulatory component; T-cell mediated immune response; DC;
 KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
 KW PAP protein; Ala Arg linker; membrane distal extracellular domain;
 KW membrane distal intracellular domain; C-terminal tag; human; GM-CSF;
 KW HER-2 protein; granulocyte-macrophage colony stimulating factor;
 KW ovalbumin-derived octapeptide; OVA; rat; HER500-rGM-CSF fusion protein.

XX OS Chimeric - Homo sapiens.
 OS Chimeric - Rattus norvegicus.
 OS Chimeric - unidentified.

PN WO200174855-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US10515.

XX 30-MAR-2000; 2000US-193504P.

XX (DEND-) DENDREON CORP.

PI Laus R, Vidovic D, Graddis T;

XX WPI; 2001-662965/76.

DR N-PSDB; AAD21567.

PT An immunostimulatory fusion protein comprising the intracellular domain
 of HER-2 and an antigen elicits an immune response to the antigen and
 is useful for the treatment of associated cancer associated -
 Claim 7; Page 27; 59pp; English.

CC The invention relates to immunostimulatory fusion proteins (IFP) and
 CC nucleic acid molecules encoding such proteins. The IFPs comprise a
 CC polypeptide antigen component and an immunostimulatory component derived
 CC from the intracellular domain of HER-2 protein which is effective to
 CC elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
 CC immune response to the antigen. IFP or superactivated dendritic cells
 CC are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
 CC associated with a particularly antigen. The present sequence is HER500
 CC rGM-CSF fusion protein construct which comprises human PAP
 CC signal sequence, mature PAP protein, an Ala Arg linker, human HER-2
 CC signal sequence, mature HER-2 membrane distal extracellular domain,
 CC an Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
 CC HER-2 membrane distal intracellular domain, an Ala Ala linker, a mature
 CC rat granulocyte-macrophage colony stimulating factor (GM-CSF) sequence
 CC and a C-terminal tag.

XX Sequence 697 AA;

Query Match 100.0%; Score 38; DB 22; Length 697;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIINEKL 8
 |||||

Db 331 SIINFEXL 338

RESULT 125

AB31611 AAB31611 standard; Protein; 948 AA.

AC AAB31611;

DT 30-APR-2001 (first entry)

DE Amino acid sequence of Hsp65-ovalbumin fusion protein.

XX Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;

KW lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;

XX ovalbumin.

OS Synthetic.

OS Mycobacterium bovis.

OS Gallus sp.

PN WO200104344-A2.

PD 18-JAN-2001.

PF 10-JUL-2000; 2000WO-US18828.

PR 08-JUL-1999; 99US-0143757.

PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

PI Siegel M, Chu NR, Mizzzen LA;

DR WPI: 2001-138361/14.

XX N-PSDB; AAF25014.

XX Screening for compounds that stimulate Th1-like responses in CD4+ T

XX lymphocyte cells -

XX Example 8; Fig 7A-B; 88pp; English.

CC The present sequence represents a fusion protein comprising a

CC Mycobacterium bovis heat shock protein (Hsp) 65 fused at its C terminal

CC to an ovalbumin protein. The fusion protein is used in the method of the

CC invention. The specification describes a method of determining whether a

CC compound stimulates a Th1-like response. Th1 cells are a subset of CD4+

CC T lymphocyte cells. The method comprises contacting naive lymphocytes

CC in vitro with a fusion protein comprising at least a fragment of Hsp,

CC and then detecting the Th1-like response exhibited by the cell sample.

CC The proteins which may be used in the method of the invention are Hsp65,

CC Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify

CC compounds that stimulate Th1-like responses in response to microbial

CC pathogens.

CC

CC

CC

SQ Sequence 948 AA;

Query Match 100.0%; Score 38; DB 22; Length 948;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8

DB 820 SIINFEXL 827

Search completed: January 30, 2004, 07:17:23
Job time : 74 secs

! FINDPATTERNS on geneseqp: * allowing 0 mismatches
! 1 C(R,K){20,20} January 30, 2004 07:26 ..

1 AAW45801 ck: 9500 len: 39 ! Aaw45801 One chain of a bombesin dimer. 6/1

C(R,K){20,20}
C(K){20}
1: C(K){20} XGGGQ

1 AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen PK: 11/200

C(R,K){20,20}
C(K){20}
1: C(K){20}

1 ABG92659 ck: 5509 len: 58 ! Abg92659 Human DNA-binding protein #85. 11/

C(R,K){20,20}
C(K){20}
35: KFYFV C(K){20}

1 AAU18238 ck: 5509 len: 58 ! Aau18238 Novel human DNA-binding protein #8

C(R,K){20,20}
C(K){20}
35: KFYFV C(K){20}

1 AA003766 ck: 8808 len: 81 ! Aa003766 Human polypeptide SEQ ID NO 17658.

C(R,K){20,20}
C(K){20}
48: LTTTA C(K){20}

1 AA011210 ck: 863 len: 70 ! Aa011210 Human polypeptide SEQ ID NO 25102.

C(R,K){20,20}
C(K){20}
30: IDLCL C(K){20}

Databases searched:
Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 6
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 04:40.14

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
! 1 C(R,K) {20,20}

January 30, 2004 07:38 ..

GENESEQP1990S:AAM45801	ck: 9500	len: 39	finds: 1	! Aaw45801 One chain of a bomb
GENESEQP2000S:AAB13780	ck: 7317	len: 21	finds: 1	! Aab13780 Soluble peptide and
GENESEQP2002S:ABG92659	ck: 5509	len: 58	finds: 1	! Abg92659 Human DNA-binding P
GENESEQP2001S:AAU18238	ck: 5509	len: 58	finds: 1	! Aau18238 Novel human DNA-bir
GENESEQP2001S:AAO03766	ck: 8808	len: 81	finds: 1	! Aao03766 Human polypeptide S
GENESEQP2001S:AAO11210	ck: 863	len: 70	finds: 1	! Aao11210 Human polypeptide S

\\End of list

Databases searched:

Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 6
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 07:02.46

PN US2002102638-A1.
 XX
 PD 01-AUG-2002.
 XX
 PF 17-JAN-2001; 2001US-0764846.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225159P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-226688P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229513P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234223P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC,
 DR WPI; 2002-690611/74.
 DR N-PSDB; ABS68254.
 XX
 PT Novel DNA-binding protein useful for diagnosis, prognosis, prevention
 PT and treatment of immune, hyperproliferative, respiratory,
 PT cardiovascular, reproductive, endocrine, gastrointestinal and
 PT neurological disorders
 PS Claim 11; SEQ ID No 223; 225bp; English.
 XX

CC The present invention relates to a new DNA-binding protein. The invention
 CC is useful in treating, preventing, diagnosing and/or prognosing
 CC immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
 CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
 CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.
 CC asthma), inflammatory conditions, graft-versus-host disease, blood-
 CC related disorders (thrombosis, atherosclerosis), hyperproliferative
 CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
 CC cardiovascular disorders (e.g. arrhythmia), respiratory disorders
 CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
 CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's
 CC disease), reproductive system disorders (e.g. endometriosis),
 CC infectious diseases (e.g. viral, bacterial or fungal infections) and
 CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also
 CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose
 CC neuronal damage which occurs in certain neuronal disorders or neuro-
 CC degenerative conditions. The present amino acid sequence represents a
 CC human DNA-binding protein of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from USPTO at
 CC http://seqdata.uspto.gov/sequence.
 XX
 SQ Sequence 58 AA;
 XX
 ABG92659 Length: 58 January 30, 2004 07:48 Type: P Check: 5509
 1 TYLECEHNSL VNSKCLTVVL SRCISVCLNK FFFVCKKKKK KKKKKKKKKK
 51 KKKKKKKK
 !!AA SEQUENCE 1.0
 ID AAU18238 standard; Protein: 58 AA.
 AC AAU18238;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Novel human DNA-binding protein #85.
 XX
 KW Human, DNA-binding protein; histone, chromo domain protein;
 KW chromatin organisation modifier; Y-box binding protein;
 KW DNA organisation; gene transcription; malignant disease;
 KW autoimmune disorder; rheumatic disease; genetic abnormality;
 KW infectious disease; neurological disorder; gene therapy;
 KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
 KW cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200155162-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01305.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0215135.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228927.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0232403.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249249.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Baraeth SC, Ruben SM;

WPI; 2001-465557/50.

DR N-PSDB; AAS29114.

Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Claim 11; SEQ ID No 223; 561pp; English.

The present invention relates to the isolation of novel DNA-binding proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromatin (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's

CC disease). The polynucleotide sequences of the invention may also be
 CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
 CC proteins.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 58 AA;

AAU18238 Length: 58 January 30, 2004 07:48 Type: P Check: 5509 ..

1 TYIECHNSL VNSKCLTVVL SRCISVCLNK FYFVCKKKKK KKKKKKKKK

51 KKKKKKKK

11AA_SEQUENCE 1.0
 ID_AA003766 standard; Protein; 81 AA.

AA003766;

06-NOV-2001 (first entry)

Human polypeptide SEQ ID NO 17658.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 Human vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 tissue growth factor; immunomodulatory; cancer; leukaemia;
 nervous system disorders; arthritis; inflammation.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US04927.

28-FEB-2000; 2000US-0515126.

18-MAY-2000; 2000US-0577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI: 2001-514838/56.

N-PSDB; AA183637.

Isolated nucleic acids and polypeptides, useful for preventing
 diagnosing and treating e.g. leukaemia, inflammation and immune
 disorders -

Claim 20; SEQ ID NO 17658; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AA179941-AA193841) and
 the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 cytokine, cell proliferation or cell differentiation or which may induce
 production of other cytokines in other cell populations. The
 polynucleotides and polypeptides are useful in gene therapy, vaccines or
 peptide therapy. The polypeptides have various cytokine-like activities,
 e.g. stem cell growth factor activity, haematopoiesis regulating
 activity, tissue growth factor activity, immunomodulatory activity and
 activin/inhibin activity and may be useful in the diagnosis and/or
 treatment of cancer, leukaemia, nervous system disorders, arthritis and
 inflammation.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 81 AA;

AA003766 Length: 81 January 30, 2004 07:48 Type: P Check: 8808 ..

1 GLNQTQLRKI LAMSSITHIG XIAVLPYNP NITILNTIY IILTTRACK

51 KKKKKKKKK KKKKKKKKK KKKKKGGG A

11AA_SEQUENCE 1.0
 ID_AA011210 standard; Protein; 70 AA.

AA011210;

06-NOV-2001 (first entry)

Human polypeptide SEQ ID NO 25102.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 tissue growth factor; immunomodulatory; cancer; leukaemia;
 nervous system disorders; arthritis; inflammation.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US04927.

28-FEB-2000; 2000US-0515126.

18-MAY-2000; 2000US-0577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI: 2001-514838/56.

N-PSDB; AA191141.

Isolated nucleic acids and polypeptides, useful for preventing
 diagnosing and treating e.g. leukaemia, inflammation and immune
 disorders -

Claim 20; SEQ ID NO 25102; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AA179941-AA193841) and
 the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 cytokine, cell proliferation or cell differentiation or which may induce
 production of other cytokines in other cell populations. The
 polynucleotides and polypeptides are useful in gene therapy, vaccines or
 peptide therapy. The polypeptides have various cytokine-like activities,
 e.g. stem cell growth factor activity, haematopoiesis regulating
 activity, tissue growth factor activity, immunomodulatory activity and
 activin/inhibin activity and may be useful in the diagnosis and/or
 treatment of cancer, leukaemia, nervous system disorders, arthritis and
 inflammation.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 70 AA;

AA011210 Length: 70 January 30, 2004 07:48 Type: P Check: 863 ..

1 YYIHRTVPM CMNXLKDNV DKXTIDLCLC KKKKKKKKK KKKKKKKKK

51 KKKKKKKKK KKKKPGGGA

! FINDPATTERNS on pir:* allowing 0 mismatches
! 1 C(R,K) {20,20}

January 30, 2004 06:58 ..

Databases searched:

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 0
Total length: 96,168,682
Total sequences: 283,308
CPU time: 01:56.42

! FINDPATTERNS on swp:* allowing 0 mismatches
! 1 C(R,K) {20,20} January 30, 2004 06:58 ..

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 0
Total length: 305,079,309
Total sequences: 958,388
CPU time: 06:50.54

> 0 <
01:10 Intelligenetics
> 0 <

Quest - Quick User-directed Expression Search Tool
Release 5.4

-- Outline of search "seq6-1ss" --

Selected search type is key against sequence data banks or files.

Selected scope is Sequence.

Selected sequence key from "new.key":

seq6 (AA) ID seq6 AA preliminary pattern

1 followed by

2 c

2 r or k repeated 20 times

Selected data banks and files:

Data bank : Issued_AA , all entries

-- Output Parameters --

Format Options:

Nucleic acid code matching

Find non-matching hits only

Report key used

Note position of hit

Display full annotations

Sequence context

-- Run Parameters --

Run mode

Time to start comparison

Notify at end of run

No hits found.

-- Search Statistics --

Times:

CPU 00:01:29.16

Total Elapsed 00:01:30.00

Number of sequences searched:

Number of sequence hits:

Number of separate matches:

Number of sequence hits saved:

File Options:
Indirect file No
Sequence or key file No
List of hits No
Hit display Yes
Name and annotations Yes

328807
0
0
0

! FINDPATTERNS on pir:* allowing 0 mismatches

! 1 (R,K){20,20}

January 30, 2004 07:05 ..

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 17
Total length: 96,168,682
Total sequences: 283,308
CPU time: 03:26.44

1 T49173 ck: 4143 len: 517 ! hypothetical protein T20N10.250 - Arabidops

(R,K){20,20}

(K){20}

444: FERVG KKKKKKKKKKKKKKKKKKK KKKIR

445: ERVGR KKKKKKKKKKKKKKKKKKK KKIRL

(K){20}

446: RVGKK KKKKKKKKKKKKKKKKKKK KIRLN

(K){20}

447: VGKKK KKKKKKKKKKKKKKKKKKK IRLNF

1 T46395 ck: 7330 len: 380 ! hypothetical protein DKFZp434I1120.1 - huma

(R,K){20,20}

(K){20}

355: NLLIQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

356: LLLQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

357: LLQKK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

358: LQKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

359: QKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

360: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

361: KKKKK KKKKKKKKKKKKKKKKKKK

1 I52523 ck: 8048 len: 215 ! nucleoporin p62 homolog - rat (fragment)

(R,K){20,20}

(K){20}

35: CEFLE KKKKKKKKKKKKKKKKKKK KKTGD

(K){20}

36: EFLEK KKKKKKKKKKKKKKKKKKK KTGDN

(K){20}

37: FLEKK KKKKKKKKKKKKKKKKKKK TGDNA

1 S58321 ck: 1384 len: 126 ! probable membrane protein YOR309c - yeast

(R,K){20,20}

(K){20}

53: KKRRT RRRRRRRRRRRRRRRRRR KRSPR

(R,K){20}

54: KRPTR RRRRRRRRRRRRRRRR RSPRK

(R,K){20}

55: RRTTR KRRRRRRRRRRRRRRR SPRKR

Databases searched:


```

!!SEQUENCE_LIST 1.0
! FINDPATTERNS on plr:* allowing 0 mismatches
!      1 (R,K){20,20}      January 30, 2004 07:51 ..

PIR2:T49173      ck: 4143 len: 517 finds: 4 ! hypothetical protein T20N10.25
PIR2:T46395      ck: 7330 len: 380 finds: 7 ! hypothetical protein DKFZp434I
PIR2:I52523      ck: 8048 len: 215 finds: 3 ! nucleoporin p62 homolog - rat
PIR2:S58321      ck: 1384 len: 126 finds: 3 ! probable membrane protein YOR3

\\End of list

Databases searched:
  NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds:      17
Total length:    96,168,682
Total sequences: 283,308
CPU time:        03:53.65

```



```

!!AA SEQUENCE 1.0
P1:T91913 - hypothetical protein T20N10.250 - Arabidopsis thaliana
C1:Species: Arabidopsis thaliana (mouse-ear cress)
C1:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C1:Accession: T91913
R1:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.;
Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quecler, F.; Salanoubat, M.
submitted to the Protein Sequence Database, April 2000
A1:Reference number: Z25017
A1:Accession: T91913
A1:Status: preliminary
A1:Molecule type: DNA
A1:Residues: 1-517 <DNA>
A1:Cross-references: EMBL:AL35032; GSPDB:GN00061; ATSP:T20N10.250
A1:Experimental source: cultivar Columbia; BAC clone T20N10
C1:Genetics:
A1:Gene: ATSP:T20N10.250
A1:Map position: 3
A1:Introns: 312/3; 359/3; 444/3
C1:Superfamily: Arabidopsis thaliana hypothetical protein P17J16.30
T91913 Length: 517 January 30, 2004 07:58 Type: P Check: 4143 ..

1 MDLFSLEPNE LLYHLSFLS TKEALTSVL SKRWNLFAF VYLEFDDSV
51 FLHPERKKE KEGILQFMD FVDRVLDLHG DSLIKTFSLK CKTVGSDHV
101 DRWICNVLAR GVSDDLDFID FRDLYSLPHE VGVSRLLVVL RVGSSDLVW
151 WQKFLCLPML KTLVLDSCML CIGQFOILL ACPALBELDM TNRWKDSNV
201 TVSSSILKEL TIDHGCCSV VNLKLSFDA PSLVIFYCD SLAEDYPQVN
251 LKNLYBAQIN LLLTQAOIEQ VRALNEMLV ADDVFPGLGN AMKLITGLRN
301 VQQLYSPDT LEVLSRCEG MPVFNNLKV LSIWSDMNRGM QAMPVLLRNC
351 PHLETLIIRG LLYHATDKCG DVCCISRDY KQHSLSFCPV KKLQIYFRG
401 TIRELEMIKH FLKIPPCKE MDIYAHNSH TLFKDTTIE RVGKKKKKKK
451 KKKKKKKKK KKKKKIRLN FRPVNKTQF LKRLADKCF IPQCLEPLDV
501 DSSLGELAIL AMDSRPS

!!AA SEQUENCE 1.0
P1:T6395 - hypothetical protein DKFZp43411120.1 - human (fragment)
C1:Species: Homo sapiens (man)
C1:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C1:Accession: T6395
R1:Ottenwajlder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A1:Reference number: Z23031
A1:Accession: T6395
A1:Status: preliminary
A1:Molecule type: mRNA
A1:Residues: 1-380 <AA>
A1:Cross-references: EMBL:AL137556
A1:Experimental source: adult testis; clone DKFZp43411120
C1:Genetics:
A1:Note: DKFZp43411120.1
T6395 Length: 380 January 30, 2004 07:58 Type: P Check: 7330 ..

1 WGSTDSKLNK RKAIVQLTTK TQPEARDDA FMDQFMA DTA TSVQDFVALV
51 PAAEIRAVRE ESPNLATLC YKAVERLVQG AEGSGHSEKE KQIVANCRL
101 LRRVLPYIRE DPMRGFFWS TVPAGRGCG EEDDEHARPL AESLLATAD
151 LIFCPDFTVQ SHRSTVDSA EDVHSLDSC YIWEAGVGR HSPQNYIHD
201 NNRMEILKLL LTCFSEAMYL PPAEESGSTN PWQFCSTN NRHALPLFTS

!!AA SEQUENCE 1.0
P1:558321 - probable membrane protein YOR309c - yeast (Saccharomyces cerevisiae)
M1:Alternate names: hypothetical protein 06105
C1:Species: Saccharomyces cerevisiae
C1:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002
C1:Accession: S58321; S67215; S71989
R1:Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
submitted to the EMBL Data Library, August 1995
A1:Reference number: S58318
A1:Accession: S58321
A1:Molecule type: DNA
A1:Residues: 1-126 <PEA>
A1:Cross-references: EMBL:X90565; NID:G940836; PID:G940840
R1:Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.
submitted to the Protein Sequence Database, July 1996
A1:Reference number: S67213
A1:Accession: S67215
A1:Molecule type: DNA
A1:Residues: 1-126 <PEW>
A1:Cross-references: EMBL:Z75217; NID:G1420680; PID:e252431; PID:G1420681;
MIPS:YOR309C
A1:Experimental source: strain S288C
R1:Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.;
Schweizer, M.
Yeast 12, 1021-1031, 1996
A1>Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast
chromosome XV reveals regions of similarity to chromosomes I and XIII.
A1:Reference number: S71989; MUID:97051589; PMID:8896266
A1:Accession: S71989
A1:Status: nucleic acid sequence not shown; translation not shown
A1:Molecule type: DNA
A1:Residues: 1-126 <PEF>
A1:Cross-references: EMBL:X90565; NID:G940836; PID:CA62164.1; PID:G940840
A1:Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1995
C1:Genetics:
A1:Cross-references: SGD:S0005836

```

A:Map position: 15R
 C:Keywords: transmembrane protein
 F:3-19/Domain: transmembrane #status predicted <TM1>
 F:107-123/Domain: transmembrane #status predicted <TM2>
 SS8321 Length: 126 January 30, 2004 07:58 Type: P Check: 1384 ..
 1 MQMLIPQRL LILNPLMMK RKKKKKKKKR RERETWKKIP RILKKLRRKR
 51 RTRRRKKKKR KRRRRKKKKR RKKRSPKKR KRRNKDAFYI LIIDPSRSL
 101 LFGFRKFSII IQCLTYFSFH ILFHNL

FINDPATTERNS on swp: * allowing 0 mismatches
1 (R,K){20,20} January 30, 2004 07:00 ..

1 Q12444 ck: 1384 len: 126 1 Q12444 saccharomyces cerevisiae (baker's ye
(R,K){20,20}
(R,K){20}
53: RKRRT RRRRRRRRRRRRRRRRRR KRSPR
(R,K){20}
54: KRRTT RRRRRRRRRRRRRRRRRR RSPRK
(R,K){20}
55: RRTTR RRRRRRRRRRRRRRRR SPRKR

1 Q9P529 ck: 291 len: 128 1 Q9P529 neurospora crassa. hypothetical 15.2
(R,K){20,20}
(K){20}

71: KRNQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
72: RKNQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
73: KNQK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
74: NQKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
75: QKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
76: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
77: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
78: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
79: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
80: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
81: KKKK KKKKKKKKKKKKKKKKKKK KKEOE
(K){20}
82: KKKK KKKKKKKKKKKKKKKKKKK KKEOS
(K){20}
83: KKKK KKKKKKKKKKKKKKKKKKK EDESR

1 Q9NT34 ck: 7330 len: 380 1 Q9nt34 homo sapiens (human). hypothetical 1
(R,K){20,20}
(K){20}
355: NLLQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
356: LLLQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
357: LLOK KKKKKKKKKKKKKKKKKKK KKKK

(K){20}
358: LQKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
359: OKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
360: KKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
361: KKKK KKKKKKKKKKKKKKKKKKK

1 Q9H6Q7 ck: 3351 len: 720 1 Q9h6q7 homo sapiens (human). hypothetical 1
(R,K){20,20}
(K){20}

692: IVSIS KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
693: VSISK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
694: SISKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
695: ISKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
696: SKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
697: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
698: KKKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
699: KKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
700: KKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
701: KKKK KKKKKKKKKKKKKKKKKKK

1 Q8N6F0 ck: 9898 len: 55 1 Q8n6f0 homo sapiens (human). similar to loc
(R,K){20,20}
(R,K){20}

14: RRGGR KKKRRRRRRRRRRRRRRRR KKKK
(R,K){20}
15: RGRG KKKRRRRRRRRRRRRRRRR KKKK
(R,K){20}
16: GRGK KKKRRRRRRRRRRRRRRRR KKKK
(R,K){20}
17: RGKK KKKRRRRRRRRRRRRRRRR KKKK
(R,K){20}
18: GKKK RRRRRRRRRRRRRRRRRRR KKKK
(R,K){20}
19: KKKR KRRRRRRRRRRRRRRRRRR KKKR
(R,K){20}
20: KKKR RRRRRRRRRRRRRRRRRRR KRRR
(K){20}
21: KRRR KKKKKKKKKKKKKKKKKKK KRRR

22:	KRKR	KKKKKKKKKKKKKKKKKK	RRRR
		(R, K) {20}	
23:	KRKX	KKKKKKKKKKKKKKKKKK	RRRR
		(R, K) {20}	
24:	KRKX	KKKKKKKKKKKKKKKKKK	RRRG
		(R, K) {20}	
25:	KKXK	KKKKKKKKKKKKKKKKRR	RRGR
		(R, K) {20}	
26:	KKXK	KKKKKKKKKKKKKKRRRR	RGRR
		(R, K) {20}	
27:	KKXK	KKKKKKKKKKKKRRRR	RGRR
		(R, K) {20}	
28:	KKXK	KKKKKKKKKKKKRRRR	GRRR
		(R, K) {20}	

```
Q9HC48 ck: 7602 len: 667 ! Q9hc48 homo sapiens (human). ctcl tumor ant  
      (R,K) {20,20}  
      (R,X) {20}  
648: GDKTDD RKKKKKKKKKKKKKKKKKK
```

```

      (K) {20}
186: KIFLK KKKKKKKKKKKKKKKKKKK T

Q95IV6 ck: 7515 len: 531 | Q95IV6 macaca fascicularis (crab eating mac

```

Q9H5V6 ck: 379 len: 168 ! Q9H5V6 homo sapiens (human) . hypothetical F

140: VREME $(R, X) \{20, 20\}$
 $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

141: REMEK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

142: EWEKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

143: WEKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

144: EKKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

145: KKKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

146: KKKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

147: KKKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

148: KKKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

149: KKKKK $(K) \{20\}$ KKKKKKKKKKKKKKKKKKKKKKKKKKK

```

1                                     (R,X) {20,20}
502: YKGN$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
503: KGN$K KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
504: GNS$K KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
505: NS$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
506: SK$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
507: K$K$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
508: K$K$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KKKK
509: K$K$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KKK
510: K$K$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} KK
511: K$K$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20} K
512: K$K$K$ KKKKKKKKKKKKKKKKKKKKK (X) {20}

```

```

1      Q8SMR7 ck: 2372 len: 515 ! Q8smr7 drosophila melanogaster (fruit fly)
      (R,K) {20,20}
      (K) {20}
493: FTDYI KKKKKKKKKKKKKKKKKKKKK KKK
      (K) {20}
494: TDYIK KKKKKKKKKKKKKKKKKKKKK KK

```

```

1      Q9LXR2  ck: 4143  len: 517   | Q9LXR2 arabidopsis thaliana (mouse-ear cress)
      (R,K){20,20}
      (K){20}
444:  FERVG  KKKKKKKKKKKKKKKKKKKKK KKKIR
      (K){20}
445:  ERVGK  KKKKKKKKKKKKKKKKKKKKK KKIRL

```


446: RVGKK (K){20}
 447: VGGKK (K){20} IRNLF

1
 08s7d3 ck: 6479 len: 80 i 08s7d3 oryza sativa (rice). hypothetical 9.

48: VIHLD (R,K){20,20}
 49: IHLDK (K){20}
 50: HLDKK (K){20}
 51: LDGKK (K){20}
 52: DKKKK (K){20}
 53: KKKKK (K){20}
 54: KKKKK (K){20}
 55: KKKKK (K){20}
 56: KKKKK (K){20} KLGE
 57: KKKKK (K){20} LGE

1
 08Lp6 ck: 3239 len: 113 i 08Lp6 oryza sativa (japonica cultivar-group

10: SLEHT (R,K){20,20}
 11: LEHIK (K){20}
 12: EHIKK (K){20}
 13: HIKKK (K){20}
 14: IKKKK (K){20}
 15: KKKKK (K){20} KRRR
 16: KKKKK (K){20} KRRR
 17: KKKKK (K){20} RRRR
 18: KKKKK (R,K){20} ERRR

Q9Lg29 ck: 6094 len: 260 i Q9Lg29 arabidopsis thaliana (mouse-ear cres

1

6: MDRCI (R,K){20,20}
 7: DRCIR (R,K){20}
 8: RCIRK (K){20}
 9: CIRKK (K){20}
 10: IRKKK (K){20}
 11: RKKKK (K){20}
 12: KKKKK (K){20}
 13: KKKKK (K){20}
 14: KKKKK (K){20}
 15: KKKKK (K){20}
 16: KKKKK (K){20}
 17: KKKKK (K){20}
 18: KKKKK (K){20}
 19: KKKKK (K){20}
 20: KKKKK (K){20}
 21: KKKKK (K){20}
 22: KKKKK (K){20}
 23: KKKKK (K){20}
 24: KKKKK (K){20}
 25: KKKKK (K){20}
 26: KKKKK (K){20}
 27: KKKKK (K){20}
 28: KKKKK (K){20}
 29: KKKKK (K){20}

78:	XXXXX	XXXXXXXXXXXXXXXXXXXXX	XXXXX
79:	XXXXX	(X) { 20 }	XXXXX
80:	XXXXX	XXXXXXXXXXXXXXXXXXXXX	XXXXX
81:	XXXXX	(X) { 20 }	XXXXX
82:	XXXXX	XXXXXXXXXXXXXXXXXXXXX	XXXXX
83:	XXXXX	(X) { 20 }	XXXXX
84:	XXXXX	XXXXXXXXXXXXXXXXXXXXX	XXXXX
85:	XXXXX	(X) { 20 }	XXXXX
86:	XXXXX	XXXXXXXXXXXXXXXXXXXXX	XXXXX
87:	XXXXX	(X) { 20 }	XXXXX
88:	XXXXX	(X) { 20 }	XXXXX
89:	XXXXX	(X) { 20 }	XXXXX
90:	XXXXX	(X) { 20 }	XXXXX
91:	XXXXX	(X) { 20 }	XXXXX
92:	XXXXX	(X) { 20 }	XXXXX
93:	XXXXX	(X) { 20 }	XXXXX
94:	XXXXX	(X) { 20 }	XXXXX
95:	XXXXX	(X) { 20 }	XXXXX
96:	XXXXX	(X) { 20 }	XXXXX
97:	XXXXX	(X) { 20 }	XXXXX
98:	XXXXX	(X) { 20 }	XXXXX
99:	XXXXX	(X) { 20 }	XXXXX
100:	XXXXX	(X) { 20 }	XXXXX
101:	XXXXX	(X) { 20 }	XXXXX
102:	XXXXX	(X) { 20 }	XXXXX

103:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
104:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
105:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
106:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
107:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
108:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
109:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
110:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
111:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
112:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
113:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
114:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
115:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
116:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
117:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
118:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
119:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
120:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
121:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
122:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
123:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
124:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
125:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
126:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	

127: KKKK (K) {20} KKKK
128: KKKK (K) {20} KKKK
129: KKKK (K) {20} KKKK
130: KKKK (K) {20} KKKK
131: KKKK (K) {20} KKKK
132: KKKK (K) {20} KKKK
133: KKKK (K) {20} KKKK
134: KKKK (K) {20} KKKK
135: KKKK (K) {20} KKKK
136: KKKK (K) {20} KKKK
137: KKKK (K) {20} KKKK
138: KKKK (K) {20} KKKK
139: KKKK (K) {20} KKKK
140: KKKK (K) {20} KKKK
141: KKKK (K) {20} KKKK
142: KKKK (K) {20} KKKK
143: KKKK (K) {20} KKKK
144: KKKK (K) {20} KKKK
145: KKKK (K) {20} KKKK
146: KKKK (K) {20} KKKK
147: KKKK (K) {20} KKKK
148: KKKK (K) {20} KKKK
149: KKKK (K) {20} KKKK
150: KKKK (K) {20} KKKK

151: KKKK (K) {20} KKKK
152: KKKK (K) {20} KKKK
153: KKKK (K) {20} KKKK
154: KKKK (K) {20} KKKK
155: KKKK (K) {20} KKKK
156: KKKK (K) {20} KKKK
157: KKKK (K) {20} KKKK
158: KKKK (K) {20} KKKK
159: KKKK (K) {20} KKKK
160: KKKK (K) {20} KKKK
161: KKKK (K) {20} KKKK
162: KKKK (K) {20} KKKK
163: KKKK (K) {20} KKKK
164: KKKK (K) {20} KKKK
165: KKKK (K) {20} KKKK
166: KKKK (K) {20} KKKK
167: KKKK (K) {20} KKKK
168: KKKK (K) {20} KKKK
169: KKKK (K) {20} KKKK
170: KKKK (K) {20} KKKK
171: KKKK (K) {20} KKKK
172: KKKK (K) {20} KKKK
173: KKKK (K) {20} KKKK
174: KKKK (K) {20} KKKK
175: KKKK (K) {20} KKKK

176: KKKK (K) {20} KKKK
177: KKKK (K) {20} KKKK
178: KKKK (K) {20} KKKK
179: KKKK (K) {20} KKKK
180: KKKK (K) {20} KKKK
181: KKKK (K) {20} KKKK
182: KKKK (K) {20} KKKK
183: KKKK (K) {20} KKKK
184: KKKK (K) {20} KKKK
185: KKKK (K) {20} KKKK
186: KKKK (K) {20} KKKK
187: KKKK (K) {20} KKKK
188: KKKK (K) {20} KKKK
189: KKKK (K) {20} KKKK
190: KKKK (K) {20} KKKK
191: KKKK (K) {20} KKKK
192: KKKK (K) {20} KKKK
193: KKKK (K) {20} KKKK
194: KKKK (K) {20} KKKK
195: KKKK (K) {20} KKKK
196: KKKK (K) {20} KKKK
197: KKKK (K) {20} KKKK
198: KKKK (K) {20} KKKK
199: KKKK (K) {20} KKKK

200: KKKK (K) {20} KKKK
201: KKKK (K) {20} KKKK
202: KKKK (K) {20} KKKK
203: KKKK (K) {20} KKKK
204: KKKK (K) {20} KKKK
205: KKKK (K) {20} KKKK
206: KKKK (K) {20} KKKK
207: KKKK (K) {20} KKKK
208: KKKK (K) {20} KKKK
209: KKKK (K) {20} KKKK
210: KKKK (K) {20} KKKK
211: KKKK (K) {20} KKKK
212: KKKK (K) {20} KKKK
213: KKKK (K) {20} KKKK
214: KKKK (K) {20} KKKK
215: KKKK (K) {20} KKKK
216: KKKK (K) {20} KKKK
217: KKKK (K) {20} KKKK
218: KKKK (K) {20} KKKK
219: KKKK (K) {20} KKKK
220: KKKK (K) {20} KKKK
221: KKKK (K) {20} KKKK
222: KKKK (K) {20} KKKK
223: KKKK (K) {20} KKKK

224: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKR
 (K) {20}
225: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKRRH
 (K) {20}
226: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KORHH
 (K) {20}
227: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK NRHHI

Q64075 ck: 8048 len: 215 ; Q64075 rattus sp. nucleoporin p62 homolog R
 (R,K) {20,20}
 (K) {20}

35: CEFLK KKKKKKKKKKKKKKKKKKKKKKKKKK KKTGD
 (K) {20}

36: EFLEK KKKKKKKKKKKKKKKKKKKKKKKKKK KTGDN
 (K) {20}

37: FLEKK KKKKKKKKKKKKKKKKKKKKKKKKKK TGDNA
 (K) {20}

Q9D5G1 ck: 9388 len: 169 ; Q9d5g1 mus musculus (mouse). adult male tes
 (R,K) {20,20}
 (R) {20}

117: VOLRG RRRRRRRRRRRRRRRRRRRRRR RRKKK
 (R) {20}

118: QLGR RRRRRRRRRRRRRRRRRRRRRR RKKEE
 (R) {20}

119: LRGRR RRRRRRRRRRRRRRRRRRRRRR KKEEE
 (R,K) {20}

120: RGRRR RRRRRRRRRRRRRRRRRRRRRR KKEEE
 (R,K) {20}

121: GRRRR RRRRRRRRRRRRRRRRRRRRRR KEEEE
 (R,K) {20}

122: RRRRR RRRRRRRRRRRRRRRRRRRRRR EEEEE

Q35807 ck: 7510 len: 129 ; O35807 rattus norvegicus (rat). microvascul
 (R,K) {20,20}
 (K) {20}

85: VLLAS KKKKKKKKKKKKKKKKKKKKKKKKKK KIKWE
 (K) {20}

86: LLASK KKKKKKKKKKKKKKKKKKKKKKKKKK IKWEG

Q8BXG9 ck: 5434 len: 115 ; Q8bxg9 mus musculus (mouse). hypothetical a
 (R,K) {20,20}
 (R,K) {20}

40: IIIII RRRRRRRRRRRRRRRRRRRRRR KETGS
 (R,K) {20}

41: IIIIR RRRRRRRRRRRRRRRRRRRR ETGSH

Q8BHV2 ck: 8958 len: 154 ; Q8bhv2 mus musculus (mouse). weakly simila
 (R,K) {20,20}
 (R) {20}

42: RGEER KRRRRRRRRRRRRRRRRRRRRR RRRRR

43 :	RGEER	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
44 :	GEEER	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
45 :	EERRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
46 :	ERRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
47 :	RRRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
48 :	RRRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
49 :	RRRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
50 :	RRRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R) {20}		
51 :	RRRRR	RRRRRRRRRRRRRRRRRRR	KRRRE
	(R) {20}		
52 :	RRRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R, K) {20}		
53 :	RRRRR	RRRRRRRRRRRRRRRRRRR	RRRRR
	(R, K) {20}		

```

Databases searched:
  SWISS-PROT, Release 41.1, Released on 6jun2003, Formatted on 9jun2003
  SPTREMBL, Release 23.0, Released on 4mar2003, Formatted on 7mar2003

Total finds: 348
Total length: 305,079,309
Total sequences: 958,388
CPU time: 11:25.82

```

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches
!      1 (R,K){20,20}

```

January 30, 2004 07:59 ..

```

SP_FUN:Q12444      ck: 1384 len: 126 finds: 3      | Q12444 saccharomyces cerevisiae
SP_FUN:Q9P529      ck: 291 len: 128 finds: 13      | Q9P529 neurospora crassa. hypc
SP_HUM:Q9NT34      ck: 7330 len: 380 finds: 7       | Q9NT34 homo sapiens (human). h
SP_HUM:Q9H6Q7      ck: 3351 len: 720 finds: 10      | Q9H6Q7 homo sapiens (human). h
SP_HUM:Q8N6F0      ck: 9898 len: 55 finds: 15      | Q8N6F0 homo sapiens (human). s
SP_HUM:Q9HC48      ck: 7602 len: 667 finds: 1       | Q9HC48 homo sapiens (human). d
SP_HUM:Q9H5V6      ck: 379 len: 168 finds: 10      | Q9H5V6 homo sapiens (human). h
SP_IN:Q8SWR7       ck: 2372 len: 515 finds: 4       | Q8SWR7 drosophila melanogaster
SP_IN:Q8T2U7       ck: 8768 len: 791 finds: 2       | Q8T2U7 dictyostelium discoideu
SP_IN:Q8I247       ck: 5951 len: 206 finds: 2       | Q8I247 plasmodium falciparum
SP_OM:Q95LV6       ck: 7515 len: 531 finds: 11      | Q95LV6 macaca fascicularis (cr
SP_PL:Q9LXR2       ck: 4143 len: 517 finds: 4       | Q9LXR2 arabidopsis thaliana (m
SP_PL:Q8S7D3       ck: 6479 len: 80 finds: 10      | Q8S7D3 oryza sativa (rice). hy
SP_PL:Q8LQP6       ck: 3239 len: 113 finds: 9       | Q8LQP6 oryza sativa (japonica
SP_PL:Q9IGZ9       ck: 6094 len: 260 finds: 222     | Q9IGZ9 arabidopsis thaliana (m
SP_RO:Q64075       ck: 8048 len: 215 finds: 3       | Q64075 rattus sp. nucleoporin
SP_RO:Q9D5G1       ck: 9388 len: 169 finds: 6       | Q9D5G1 mus musculus (mouse). a
SP_RO:Q35807       ck: 7510 len: 129 finds: 2       | Q35807 rattus norvegicus (rat)
SP_RO:Q8BXG9       ck: 5434 len: 115 finds: 2       | Q8BXG9 mus musculus (mouse). h
SP_RO:Q8BHV2       ck: 8958 len: 154 finds: 12      | Q8BHV2 mus musculus (mouse). w

```

\\End of list

Databases searched:

SWISS-PROT, Release 41.1, Released on 6jun2003, Formatted on 9jun2003
 SPTREMBL, Release 23.0, Released on 4mar2003, Formatted on 7mar2003

Total finds: 348
 Total length: 305,079,309
 Total sequences: 958,388
 CPU time: 12:54.92


```

!!AA_SEQUENCE 1.0
ID_Q12444 PRELIMINARY; PRT; 126 AA.
AC_Q12444;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE ORF YOR309C.
GN YOR309C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearson B.M., Hernando Y., Kalogeropoulos A., Schweizer M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FY1673;
RA Pearson B.M., Hernando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; 275217; CA99629.1; -.
DR EMBL; X90565; CA62164.1; -.
DR GCP; S0005836; YOR309C.
SQ SEQUENCE 126 AA; 16294 MW; 46E1F4C66480208 CRC64;

Q12444 Length: 126 January 30, 2004 08:18 Type: P Check: 1384 ..

1 MGLMPLPQRL LILNPLMMK RKKKKKKKK REREYMKIP RILKLRKKR
51 RTRRRRRRRR KRRRRKKRRR RRRKSPRRR KRRNDAFYI LIIDPSRSL
101 LFGFRKFSII IQCLTFSSFH ILFHNIL

!!AA_SEQUENCE 1.0
ID_Q9P529 PRELIMINARY; PRT; 128 AA.
AC_Q9P529;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical 15.2 kDa protein.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohenel J., Brandt P., Fartmann B., Holland R.,
RA Nykatura G., Mewes H.W., Manhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356815; CAB92638.2; -.
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DBF0765 CRC64;

Q9P529 Length: 128 January 30, 2004 08:18 Type: P Check: 291 ..

1 MAISIGLH KNINRRAPGH SVYSKNSYD PQQYDATOHY LPSQGFKAIP
51 DLITGKGKGC LSTHDKRNG KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KKEQSRITF QOHROADGIC PTPWHTTR

!!AA_SEQUENCE 1.0
ID_Q9NT34 PRELIMINARY; PRT; 380 AA.

```

```

AC_Q9NT34;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP4341120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oltewelder B., Obermaier B., Mewes H.W., Gaassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137556; CAB70810.1; -.
DR Genew; HGNC:15736; C17orf28.
KW Hypothetical protein.
FT NON TER 380
SQ SEQUENCE 380 AA; 42689 MW; 67F50DD10346AFB CRC64;

Q9NT34 Length: 380 January 30, 2004 08:18 Type: P Check: 7330 ..

1 MGSTDSKINF RKAVIQUTTK TQPEATDPA FWDQFWADTA TSVQDVFAIV
51 PAEIRAVRE ESPENLALIC YKAVELYQG AEGGCHSEK KQIVLNCRL
101 LTRVLPYTFE DPMWRGFFWS TVPGAGRGG EEDDEHARL AESLILAIAD
151 LIFCPDFTVQ SHRRSTVDSA EDVHSDSCB YIWEAGVFA HSPQPNYIHD
201 MNRMLKLKL LTCFSEAMYL PPAPESGSTN PWVQFPGSTE NRHLLPLFTS
251 LINTVCAYDP VGIGIPYNNH LFSDYREPLV EEAQVLIYT LDHDSASAS
301 PTVDDTTTGT AMDADDPGP ENLFVNYLSR IHREDFQFI LKGIARLLSN
351 LLLQKKKKKK KKKKKKKKK KKKKKKKKK

!!AA_SEQUENCE 1.0
ID_Q9H6Q7 PRELIMINARY; PRT; 720 AA.
AC_Q9H6Q7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J.T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Oktani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025632; BAB15196.1; -.
KW Hypothetical protein.
FT NON TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Q9H6Q7 Length: 720 January 30, 2004 08:18 Type: P Check: 3351 ..

1 MLTEQVEQYT KEMKNTCII EDLKNELQRN KGASTLSQOT HMKIQSTLDI
51 LKERTKEAR TAEIABADAR EKDKELVEAL KRLKDYESGV YGLDAVVEI
101 KNCKQIKIR DREIELTKE INKLELKISD FLDEBAIRE RVGLEPKTMI
151 DLTEFRNSKH LKQOQYRAEN QILKEIESL EERLDLKKK IRQWAQERGK
201 RSATSGLTTE DLNLTENISQ GDRISERKLD LLSLKNMSEA QSKNEFLSRE

```

```

251 LIEKERDLER SRTYIAFQK KLKELYENK QLEEGKEIT QAIKEMQKP
301 DVKGGETSLI IPSLERLVNA IESNNAEGIF DASLHLKAOV DQLTORNEEL
351 ROELRESRKE AINYSOOLAK ANLKIDHLEK ETSILRQSEB SNVVEKGIDL
401 PGDIAPSSAS IINSQNEYLI HLLQELEENKE KKLKNELESL EDYNNKFAVI
451 RHQOSLIYKE YLSEKETWKT ESKITKEER KLEDVQODA IKVKEYNMLL
501 NALQWDSDEM KKIILAENSRK ITVLQVNEKS LIRQYTTLVE LERQURKENE
551 KQKNELLSME AEVECKIGCL QRFKEMAFK IALQKVON SVSSELELA
601 NKQYNEITAK YRDILQKDNM LVQRTSNLEH LECENISLKE QVESINKLE
651 ITREKLHTIE QAWBOETKLG NESSMDRACK SITNSDIVSI SKKKKKKKKK
701 KKKKKKKKKK KKKKKKKKKK

!!AA_SEQUENCE 1.0
ID_Q8N6F0 PRELIMINARY; PRT; 55 AA.
AC_Q8N6F0;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to LOC201361.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC030525; AAH30525.1; -
SQ SEQUENCE 55 AA; 7251 MW; 0906032B284006BA CRC64;

Q8N6F0 Length: 55 January 30, 2004 08:18 Type: P Check: 9898 ..

1 MRLASQRERR GRGKKKKRKR KKKKKKKKKK KKKKKKKKKK KRRRRRGRGR
51 RMQOQ

!!AA_SEQUENCE 1.0
ID_Q9HC48 PRELIMINARY; PRT; 667 AA.
AC_Q9HC48;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CTCL tumor antigen se2-5 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RL MEDLINE=21143360; PubMed=1149944;
RA Eichmuller S., Ueener D., Dummer R., Stein A., Thiel D.,
RA Schendendorf D.;
RT "Serological detection of cutaneous T-cell lymphoma-associated
RT antigens.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
DR EMBL; AF177228; AAG33676.1; -
DR HSSP; Q12923; 3PDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SMO0228; PDZ; 2.
DR PROSITE; PSS0106; PDZ; 2.

```

```

FT NON TER 1 1
FT NON TER 667 667
SQ SEQUENCE 667 AA; 73499 MW; C653BC16802BAE02 CRC64;

Q9HC48 Length: 667 January 30, 2004 08:18 Type: P Check: 7602 ..

1 EHENLFREND CIVRINDDL RNRREFQAOH MFRQAMRTPI IMFHVPAAN
51 KEQVEQLSQS EKNNYSSRF SPDQYIDNR SVNSAGLHTV QBARLHNHP
101 EQLDSHSLRP HSAHPSGKPP SAPASAPQNV FSTTVSGYN TKIKGKRLNI
151 QLKGTGELE FSITSRDVTI GGSAPIYKN ILPRGAQID GRLXAGRLI
201 EVNGVDLVGK SQEEVSLR STMEGTVSL LVFRQEDAFH PRELNAPSQ
251 MQIPREYAE DEDIVLTPDG TREFLTPEVP LNDSSAGLG VSVGNRSKE
301 NHADLGFVK SIINGAASK DGRLRVNDQL IAVNGESLLG KTNQDAMETL
351 RRMSTEGNK RGMQLIVAR RISKNEKLS PGSPGPPELP IETALDRER
401 RISHSLYSGI EGLDESPSRN AALSRIWGS GKYLSPVTN MPQDDTYIIE
451 DDLRLVLPFH LSDQSSSSH DDVGFYTADA GTWAKAASD SADCSLSPDV
501 DPLVAFQREG FGRLADETK LNTVDDQXAG SPSRVDGPSL GLKXSSSLES
551 LOTVAEVTL NGDIPFHRPR PRIIRGCGN ESFRAIDKS YDKRAVDDDD
601 EGMETLEBDT EESSRSRGRES VSTASDQPSH SLERQWNGNQ EKGDKTDRKK
651 KKKKKKKKKK KKKKKKKK

!!AA_SEQUENCE 1.0
ID_Q9H5V6 PRELIMINARY; PRT; 168 AA.
AC_Q9H5V6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ22976 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tsahiro H., Ota T.,
RA Suzuki Y., Odayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isegai T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK026629; BAB15513.1; -
KW Hypothetical protein.
FT NON TER 168 168
SQ SEQUENCE 168 AA; 19549 MW; A19BDB195F8A1A90 CRC64;

Q9H5V6 Length: 168 January 30, 2004 08:18 Type: P Check: 379 ..

1 MNGNGRSGIQ QGKQNVGVA ATPTAASAC QYRCIECNOE AKELYRDYNH
51 GVLKTTICKS CQKPVDKYIE YDPIILINA ILCKAQAYRH ILFTVQINIH
101 GKLCTFLCLC EAYLRWQLO DSNQNTAPPD LIRYRENEK KKKKKKKKKK
151 KKKKKKKKKK KKKKKKKK

!!AA_SEQUENCE 1.0
ID_Q8SWR7 PRELIMINARY; PRT; 515 AA.
AC_Q8SWR7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

```

```

DE GH22607p (Fragment).
GN CG7180.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Change M., Chavez C., Dorsett V., Dreenek D., Fartin D., Frise E.,
RA George R., Gonzalez M., Guatin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celinker S.;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY095518; AAM1251.1; -.
DR Flybase; FBgn003673; CG7180.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SM00194; FTYC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KM Hydrolase.
KW
FT NON TER 515 515
SQ SEQUENCE 515 AA; 59080 MW; B2825B7E9A961958 CRC64;

Q8SWR7 Length: 515 January 30, 2004 08:18 Type: P Check: 2372 ..

1 MWTTQQLVGC PEALNEEKS PAASAAGA GADMTAATGG GSSGAGGK
51 GGRSRSSAR YDDVEKQQR SRALVSPNT IKLMLNSGL LSPERIKLEA
101 RDENSLSKT IPNGPIDRH FLKLDLRRK PVLTKLEFQ TAAYESNTC
151 RLAKKNLE KNQPKCIPY DYNRVLEKV GGLQSDSYN ASYVDSLAKP
201 NAVITQGV EETVOAYWRM WQENISAIV MLTKTFPAK VMCHQWPN
251 MEVHQQYGD I FINIVREQL ANFHITPRL YKMEKEQVT DERLILQHY
301 TEWYSHSCPF SNALLEFRRR VRLVVGNIK DEDDMGPII VHCSSGGRS
351 GYVMSIDANL ELAEEBCFN VFGYKLRQ SRKGLVENVE QYKLYDTLE
401 EHIGKTFW PVSELSDRK AKARRNSGK MNEYQAEDQ ICKQTPFTI
451 GDCAGHRAD NREKNRDLV VPDPNRPYL TSGQNAFTD YIKKKKKKKK
501 KKKKKKKKK KKKKK

```

```

DR InterPro; IPR005033; YEATS.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF03366; YEATS; 1.
DR SMART; SM00355; ZnF_C2H2; 1.
KM Hypothetical protein.
SQ SEQUENCE 791 AA; 92375 MW; D6CCB6DEC92352C CRC64;

Q8T2U7 Length: 791 January 30, 2004 08:18 Type: P Check: 8768 ..

1 MEKETETLL DTIDEKENS TTTTGNTH NNNNTNTNN NNNNNNNN
51 NNNNNNNNN NNNNNNNNN NNPSTLSAS LSKLRKLLI NNEGDRLEK
101 KRLTIMERE PPAELKKEQ ELIEIDRLN TYKEMLSLN KQNSTKST
151 YHIFTNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
201 NKKPRIFHK LQNGEPVLY CKYCRSDPV SGLGLNHR IKHGFYSTL
251 DEARQIGVP VADSEIPKD PSREGIVRF KRGAVSPSA FERNSEEDNG
301 TPDNNNNNN NNNNNNNNN NNNNNNTGN DTDKKNNG NDALVDIDL
351 NVNSYNNK EYNESSGSS RPYVKKIIV GNTSTQHPD YRGHDSYHK
401 WTVYRGPN EADISYFVK IPEYLHSPA PNDKVEYER PFNLTRGWS
451 EPPVRIPLF HDKKNRPDI IHNKLIQLP IQYVPAVVG ETTETIDLR
501 LFFDKRQQL KLDQNNNN NNNNSNNNN NNNNNNNNN NNINNNNN
551 INNNNDNN SSNTSPTSN YLNEPKIVN NDKVSENSN NEDSQKXK
601 EKEKEKEK EKEKEKEK EKEKDKXK EKEKDKXK VKEDNKEKE
651 LNNRPKDR DRDREKERK KKGKGDIE IEIEITDID IEIEIGETE
701 TETETGKE KEKEKEKE KEKEKEKG KEIEIGERG KKGKKEIE
751 TEMEIGKIE AEIEIEKEK KKKKKKKKK KKKKKKKKE I

11AA SEQUENCE 1.0
ID AC 081247 PRELIMINARY; PRT; 206 AA.
AC 081247;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFA0475C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=2255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark C., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Fretwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hanger H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhrou A.,
RA Knights A., Konfortov B., Kyes S., LaRue N., Lawson D., Leonard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL031745; CAD49055.1; -.

```

```

KW Hypothetical protein.
SQ SEQUENCE 206 AA; 25047 MW; 1192E49A3DC4523F CRC64;
081247 Length: 206 January 30, 2004 08:18 Type: P Check: 5951 ..

1 MEGQHEKNT KIKSKKPLV VSNRKPFPNV EKSKAKPLV RDPSPDSG
51 SPNAPFPNA YKPLYDSREQ EKKIEKKLK SKNITQEKD ELKKKYN DYK
101 STDILKKKE EERLKAEVL KQEKONILTK NKKPYYSR KIKKIYQEKL
151 SSYSLKAVI KKEKTLQKE RKNIKETPK KIFLKKKKK KKKKKKKKK
201 KKKKK

11AA SEQUENCE 1.0
ID Q9SLV6 PRELIMINARY; PRT; 531 AA.
AC Q9SLV6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical 61.4 kDa protein (Fragment). (Cynomolgus monkey).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecidae; Macaca.
OC NCBI_Taxid=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB071085; BAB64479.1; -.
KW Hypothetical protein.
FT NON TER 531 531
SQ SEQUENCE 531 AA; 61389 MW; B5596B4F5CDD60C CRC64;
Q9SLV6 Length: 531 January 30, 2004 08:18 Type: P Check: 7515 ..

1 MESESSNAN MNVQHEREDK NIGMLPEPV PCYSQHLSPS TYQMDPDPC
51 KSRSEPKSE GRSSWNLSTI VQKTEQETHF RESVLEPIG YMKQSPHMQ
101 EGICVGVGK TSFPTGKSE IGSMPHDPW DENPRKWD SISEKTANP
151 KNLQTVLKL DSSIMSESEY ESRSYTLEFI GKKSITSPGH VTLKTKQLPI
201 SOLPPIRCS TENNRKKQH CFKYMKGRO WYTSIGEAR SATEYAKSP
251 SKSMIDKLF NTAAGTILSN RTHQNVYGH TTEEKEVGE NVAASSLGPL
301 DFFMPLSDS KNQNTIRLS ERKTIINPKC LTMKEKSP I SQIRKINHF
351 TTKHKKKLES NLKTKKAMW QGENVDTFR NITSFPDPS DIKQSRKQT
401 EIDMRISGLS HTQPTQIESL AEGIARCSK RRTSNLVKGT KLHDESEGE
451 KOEHLTGMP FYAENFMTNT HLRKPHLGK SEDVLLGEPF ISKQFYKGN
501 SKKKKKKKK KKKKKKKKK KKKKKKKKK K

11AA SEQUENCE 1.0
ID Q9LXR2 PRELIMINARY; PRT; 517 AA.
AC Q9LXR2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical 59.7 kDa protein.
GN T20N10.250.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidops.
OK NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rued S., Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL353032; CAB88307.1; -.
DR InterPro; IPR01810; F-box.
DR InterPro; IPR06566; FBD.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00579; FBD; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PS50181; FBOX; 1.
KW Hypothetical protein.
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;
Q9LXR2 Length: 517 January 30, 2004 08:18 Type: P Check: 4143 ..

1 MDLPSLENE LVHILSFLS TKEALTSVL SKRWNLPAF VPYLEPDSV
51 FLHPEERKRE KEGILQSFMD FVDRLVDLHG DSLITSELK CKTVSDSDHV
101 DRWICNVLAR GVSDDLDFID FDLVSLPHE VGSRTLVVL RVGSESDLYW
151 WQFLCLPML KTLVLDSCWL CIGQFILL ACPALDELDM TNRWKDSNV
201 TVSSSILKEL TIDLHGCSV VNLKLSFDA PSIVFYICD SLADYQVN
251 LKNLVEAQIN LLLTQAIQEQ VVALNEMLV ADVPEGLGN AKWLITGLRN
301 VQVLYSPDT LEVLSRCCEG MPVFNNLVYL SIWSDMNGW QAMPVLLRNC
351 PHELTILIEG LHVAYDKCG DVCDCISRDY KDHSLTSPV KKLQIYFERG
401 TIRELEMIKH FLKIFPLKE MDIYAHENSH TLFKQPTFE RVGKKKKKK
451 KKKKKKKKK KKKKKKIRLN FKPVNKTQEF LKRLADKLCF IPQCLEFLDV
501 DSSLGELLAL AMSDRPS

11AA SEQUENCE 1.0
ID Q8S7D3 PRELIMINARY; PRT; 80 AA.
AC Q8S7D3;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical 9.4 kDa protein.
GN OSJNBA0057L21.23.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OK NCBI_Taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buehl C.R., Yuan Q., Ouyang S., Liu J., Mofat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tsilintin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA Vanaberg S.E., Utterback T.R., Feldlyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBA0057L21 genomic sequence."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC087599; AAL79706.1; -.
DR Gramene; Q8S7D3; -.
KW Hypothetical protein.

```

```

SQ SEQUENCE 80 AA; 9362 MW; 0177C863133B21D8 CRC64;
0857D3 Length: 80 January 30, 2004 08:18 Type: P Check: 6479
1 MAVTNRQPIVG RRMSEAAGGA TPGSRMGHWW WPATAITRFV LSVIHLDDKK
51 KKKKKKKKKK KKKKKKKKKK KKKKKKLLGB
11AA_SEQUENCE 1.0
ID 08LOP6 PRELIMINARY; PRT; 113 AA.
AC 08LOP6;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE OJ117_G01.13 protein.
GN OJ117_G01.13
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:OJ117_G01.13";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003374; BAB9330.1; -.
DR Gramene; 08LOP6; -.
SQ SEQUENCE 113 AA; 13660 MW; 597DB0EDBE2AA3EF CRC64;
08LOP6 Length: 113 January 30, 2004 08:18 Type: P Check: 3239
1 MATSISEHIK KKKKKKKKKK KKKKKKKKKK REEBDEEBEE
51 EEBELKLNKY IWDIYEAKG INEKLMPGIY VYLAHDECY RLRLTGVDVK
101 HEAMEICLFP VOV
11AA_SEQUENCE 1.0
ID 09LG29 PRELIMINARY; PRT; 260 AA.
AC 09LG29;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone:PID9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002460; BAA97098.1; -.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;
09LG29 Length: 260 January 30, 2004 08:18 Type: P Check: 6094
1 MDRCIIRKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK
51 KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK
101 KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK
151 KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK

```

```

201 KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK
251 INWVGFIYF
11AA_SEQUENCE 1.0
ID 064075 PRELIMINARY; PRT; 215 AA.
AC 064075;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Nucleoporin p62 homolog protein (fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95151924; PubMed=7849178;
RA Wang Z.Q., Akmal K.M., Kim K.H.;
RT "An unusual nucleoporin-related messenger ribonucleic acid is present
in the germ cells of rat testis.";
RL Biol. Reprod. 51:1022-1030(1994).
DR EMBL; S75997; AAB3384.1; -.
KT Porin.
FT NON TER
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;
064075 Length: 215 January 30, 2004 08:18 Type: P Check: 8048
1 SGRATSSGD EDCUSSLPF SLSGPYKDC EFLERKKKKK KKKKKKKKKK
51 KKKKKKTGDN AKSVSRQYSL KYTKLEHAE QAKVELDFIL SQKLEEDLL
101 SPLESYVEQ SGTIYLOHAD EEREKTYKLA ENIDQOLKRM AODKDIIEH
151 IMMAGPADT SDPLQIQICKI LNAHMSLOW VDQSSALLQR RVEASRVCE
201 SRKQERSL RIAPD
11AA_SEQUENCE 1.0
ID 09D5G1 PRELIMINARY; PRT; 169 AA.
AC 09D5G1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Adult male testis cDNA, RIKEN full-length enriched library,
clone:4930444p10 product:hypothetical Arginine-rich region containing
protein, full insert sequence (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The Riken Genome Exploration Research Group Phase I & II Team;

```

```

RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RL "Functional annotation of a full-length mouse cDNA collection."
RN Nature 403:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Komoto H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Komoto H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,
RA Suni N., Ishi Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanuki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
DR EMBL; AK015384; BAB29822.2; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 169 AA; 19305 MW; 91B9959380A694CC CRC64;

Q9D5G1 Length: 169 January 30, 2004 08:18 Type: P Check: 9388 ..

1 RCTGQAGPOL RALAGPWPR LAPALLSGR ARNIAGLPAA KHARDEGAS
51 AARLPAPAPH RGGQPGDAAS LSRELASTHG RRLPHCPPL PLAQTSSLVP
101 WYRLKKPIS SVQLRGRRR RRRRRRRR RRRRRRRR RRRRRRRR RRRRRRRR RRRRRRRR
151 FLDSLEPRL TSNLHRRH

!!AA_SEQUENCE 1.0
ID Q35807; PRELIMINARY; PRT; 129 AA.
AC Q35807;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE MICOVASCULAR endothelial differentiation protein 2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=9817270E; PubMed=9511718;
RA Proels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CECs, a novel mRNA

```

```

RT species, during in vitro angiogenesis."
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL; Y08769; CAA70022.1; -.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; Kinase_1.
DR Prodom; PD000001; Prot_kinase_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM, 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;

Q35807 Length: 129 January 30, 2004 08:18 Type: P Check: 7510 ..

1 MLKPHIVEL LETYSSDGL YWFFPMGA DLCEIVKRA DAGFVSEAV
51 ASHWROGLE ALRYCHDNNI HRDVKPCV LLSKKKKKK KKKKKKKKK
101 KKKKKIKWEG RDAFWAIPV KSGGVITQ

!!AA_SEQUENCE 1.0
ID Q8BXG9; PRELIMINARY; PRT; 115 AA.
AC Q8BXG9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical arginine-rich region containing protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK047167; BAC32979.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 115 AA; 13938 MW; 6C0F7EBD669CF5 CRC64;

Q8BXG9 Length: 115 January 30, 2004 08:18 Type: P Check: 5434 ..

1 GGRGISRFE VSQYTEKPC LKPKKKKKK IIIIIIIIR RRRRRRRR
51 RRRRRRRR ETGSHFVALA SLEHLPPC WNLRYRLPH TQIGFNAIKI
101 SIKDTSLSL AFVK

!!AA_SEQUENCE 1.0
ID Q8BHV2; PRELIMINARY; PRT; 154 AA.
AC Q8BHV2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Weakly similar to hypothetical 10.3 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK082253; BAC38447.1; -.
KW Hypothetical protein.

```

SO SEQUENCE 154 AA; 19596 MW; ABE8158A4839A216 CRC64;
QBHV2 Length: 154 January 30, 2004 08:18 Type: P Check: 8958 ..
1 MVQSRKQSN VGQTSRRRN VFKEVRFSSR VPRDRERGG EEEEEEEEE
51 RRRRRRRRR RRRRRRRRR KRERERESEH EVSNANKDFG LMIHFQCAC
101 DEMGRIFISH QKRNMAFLP SGDIIDRYLS YNWVMSLPS ILAYYMLKHC
151 GGCT

1

(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

AAW38841 ck: 5252 len: 30 ! Aaw38841 Delivery peptide used in peptide m

1

(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

AAW38842 ck: 7590 len: 31 ! Aaw38842 Delivery peptide used in peptide m

1

(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

AAW38877 ck: 1129 len: 23 ! Aaw38877 Delivery peptide used in peptide m

1

(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

AAW38843 ck: 3 len: 32 ! Aaw38843 Delivery peptide used in peptide m

1

(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: KKKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

AAW38878 ck: 2949 len: 24 ! Aaw38878 Delivery peptide used in peptide m

1

(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20} KVTK
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20} VTK

AAW38844 ck: 2491 len: 33 ! Aaw38844 Delivery peptide used in peptide m

```

1      (R,K){20,20}
1:     KKKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
2:     K KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
3:     KK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
4:     KKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
5:     KKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
6:     KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
7:     KKKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
8:     KKKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
9:     KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
10:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
11:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
12:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}

AAW38679 ck: 4844 len: 25 ! Aaw38679 Delivery peptide used in peptide m
1      (R,K){20,20}
1:     KKKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
2:     K KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
3:     KK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
4:     KKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
5:     KKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
6:     KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
7:     KKKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
8:     KKKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
9:     KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
10:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
11:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
12:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}

AAW38845 ck: 5054 len: 34 ! Aaw38845 Delivery peptide used in peptide m
1      (R,K){20,20}
1:     KKKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
2:     K KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
3:     KK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
4:     KKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
5:     KKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
6:     KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
7:     KKKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
8:     KKKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
9:     KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
10:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
11:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
12:    KKKKK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}

```

```

7: KKKK (K) {20}
8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}
11: KKKK (K) {20}
12: KKKK (K) {20}
13: KKKK (K) {20}

AAW3886 ck: 6814 len: 26 ! Aaw3886 Delivery peptide used in peptide m-
1: KKKK (R,K) {20,20}
2: K KKKK (K) {20}
3: KK KKKK (K) {20}
4: KKK KKKK (K) {20}

AAW3846 ck: 7692 len: 35 ! Aaw3846 Delivery peptide used in peptide m-
1: KKKK (R,K) {20,20}
2: K KKKK (K) {20}
3: KK KKKK (K) {20}
4: KKK KKKK (K) {20}
5: KKKK (K) {20}
6: KKKK (K) {20}
7: KKKK (K) {20}
8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}
11: KKKK (K) {20}

```

12: KKKKK (K) {20} KKKKK
13: KKKKK (K) {20} KKK
14: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKK

```

1
AAW38833 ck: 9248 len: 22 | Aaw38833 Delivery peptide used in peptide m
(R,K) {20,20}
(K) {20}
XXXXXXXXXXXXXXXXXXXX XX
1:

```

```

1      AA#38834   ck: 986    len: 23    ! Aa#38834 Delivery peptide used in peptide n
          (R,K) {20,20}
          (K) {20}
1:      KKKKKKKKKKKKKKKKKKKKKKKKKKK KKK
          (K) {20}
2:      K KKKKKKKKKKKKKKKKKKKKKKKKK KK

```

```

1
AA*38835 ck: 2799 len: 24 | Aa*38835 Delivery peptide used in peptide n
      (R,K){20,20}
      (K){20}
1:      KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
2:      K KKKKKKKKKKKKKKKKKKKKKKKKK KKK
      (K){20}
3:      KK KKKKKKKKKKKKKKKKKKKKKKKKK KK

```

```

1
AAW38836 ck: 4687 len: 25 | Aaw38836 Delivery peptide used in peptide n
      (R,K){20,20}
      (K){20}
1:      KKKKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
2:      K KKKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
3:      KK KKKKKKKKKKKKKKKKKKKKKKKK
      (K){20}
4:      KKK KKKKKKKKKKKKKKKKKKKKKKKK

```

[illegible]

```
5: KKKK KKKKKKKKKKKKKKKKKKKKK KK
AAW38838 ck: 8688 len: 27 ! Aaw38838 delivery peptide used in peptide m
```

```

1:                                     (R,X) {20,20}
                                     (X) {20}
1:  KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKKK
2:  K KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKKK
   (X) {20}
3:  KK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKKK
   (X) {20}
4:  KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKK
   (X) {20}
5:  KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKK
   (X) {20}
6:  KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KK
   (X) {20}

```

```

1
AAW38796 ck: 9227 len: 22 ! Aaw38796 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}
1:
XXXXXXXXXXXXXXXXXXXXX WK

```

```

1:
  (R, K) {20, 20}
    (K) {20}
      KKKKKKKKKKKKKKKKKKKKK KWK
1:
  (K) {20}
    KKKKKKKKKKKKKKKKKKKKK KWK
2:
  K KKKKKKKKKKKKKKKKKKKKK KWK

```

```

1
AAW38798 ck: 2776 len: 24 1 Aaw38798 Delivery peptide ma
(R, K) {20, 20}
(K) {20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK WK

```

[illegible]

```
AAW38800  ck: 6625  len: 26  ! Aaw38800 Delivery peptide used in peptide m:
```

1
 1: (R,K){20,20}
 (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 2: K (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 3: KK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKK
 4: KKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKK
 5: KKKK (K){20}
 KKKKKKKKKKKKKKKKKKK WK

AAW38801 ck: 8662 len: 27 ! Aaw38801 Delivery peptide used in peptide m
 1: (R,K){20,20}
 (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 2: K (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 3: KK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 4: KKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKK
 5: KKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKK
 6: KKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK WK

AAW38802 ck: 774 len: 28 ! Aaw38802 Delivery peptide used in peptide m
 1: (R,K){20,20}
 (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 2: K (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 3: KK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 4: KKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 5: KKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKK
 6: KKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKK
 7: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK WK

AAW38803 ck: 2961 len: 29 ! Aaw38803 Delivery peptide used in peptide m
 1: (R,K){20,20}
 (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 2: K (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 4: KKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 5: KKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 6: KKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 7: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKK
 8: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK WK

AAW38804 ck: 5223 len: 30 ! Aaw38804 Delivery peptide used in peptide m
 1: (R,K){20,20}
 (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 2: K (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 3: KK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 4: KKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 5: KKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 6: KKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 7: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 8: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKK
 9: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK WK

AAW38805 ck: 7560 len: 31 ! Aaw38805 Delivery peptide used in peptide m
 1: (R,K){20,20}
 (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 2: K (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 3: KK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 4: KKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 5: KKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 6: KKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK
 7: KKKKKK (K){20}
 KKKKKKKKKKKKKKKKKKK KKKKK

8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}

AAW38806 ck: 9972 len: 32 1 Aaw38806 Delivery peptide used in peptide m

1: (R,K) {20,20}
2: K (K) {20}
3: KK (K) {20}
4: KK (K) {20}
5: KKK (K) {20}
6: KKKK (K) {20}
7: KKKK (K) {20}
8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}
11: KKKK (K) {20}

AAW38807 ck: 2459 len: 33 1 Aaw38807 Delivery peptide used in peptide m

1: (R,K) {20,20}
2: K (K) {20}
3: KK (K) {20}
4: KK (K) {20}
5: KKK (K) {20}
6: KKKK (K) {20}
7: KKKK (K) {20}

8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}
11: KKKK (K) {20}
12: KKKK (K) {20}

AAW38808 ck: 5021 len: 34 1 Aaw38808 Delivery peptide used in peptide me

1: (R,K) {20,20}
2: K (K) {20}
3: KK (K) {20}
4: KKK (K) {20}
5: KKKK (K) {20}
6: KKKK (K) {20}
7: KKKK (K) {20}
8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}
11: KKKK (K) {20}
12: KKKK (K) {20}
13: KKKK (K) {20}

AAW38881 ck: 8859 len: 27 1 Aaw38881 Delivery peptide used in peptide me

1: (R,K) {20,20}
2: K (K) {20}
3: KK (K) {20}
4: KKK (K) {20}

5: KKKK (K) {20}
KKKKKKKKKKKKKKKKKK VTK

AAW38847 ck: 405 len: 36 ! Aaw38847 Delivery peptide used in peptide m

1 (R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKK KKKKK

2: K (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKK

15: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK XK

AAW38882 ck: 979 len: 28 ! Aaw38882 Delivery peptide used in peptide m

1 (R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKK KKKKK

2: K (K) {20}
KKKKKKKKKKKKKKKKKK KKKKKV

3: KK (K) {20}
KKKKKKKKKKKKKKKKKK KKKVT

4: KKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKVTK

5: KKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKVTK

6: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKVTK

1

AAW38848 ck: 3193 len: 37 ! Aaw38848 Delivery peptide used in peptide m

(R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKK KKKKK

2: K (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

15: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKK

16: KKKKKK (K) {20}
KKKKKKKKKKKKKKKKKK XK

AAW38883 ck: 3174 len: 29 ! Aaw38883 Delivery peptide used in peptide m

1 (R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKK KKKKK

2: K (K) {20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K) {20}
KKKKKKKKKKKKKKKKKK KKKKKV

4: KKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKVT

5: KKKK (K) {20}
KKKKKKKKKKKKKKKKKK KKKVTK

(K) {20}

6: KKKKK KKKKKKKKKKKKKKKKK KVTK
(K) {20}
7: KKKKK KKKKKKKKKKKKKKKKK VTK

AAW38849 ck: 6056 len: 38 1 Aaw38849 Delivery peptide used in peptide m

1 (R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAW38884 ck: 5444 len: 30 1 Aaw38884 Delivery peptide used in peptide m

1 (R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

1

4: KKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAW38850 ck: 8994 len: 39 1 Aaw38850 Delivery peptide used in peptide m

(R,K) {20,20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK (K) {20}


```
1 AAW3885 ck: 7789 len: 31 ! Aaw3885 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK
(K){20} VTK

AAW3851 ck: 2007 len: 40 ! Aaw3851 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
```

```
14: KKKKK (K){20} KKKKK
(K){20}
15: KKKKK (K){20} KKKKK
(K){20}
16: KKKKK (K){20} KKKKK
(K){20}
17: KKKKK (K){20} KKKKK
(K){20}
18: KKKKK (K){20} KKK
(K){20}
19: KKKKK (K){20} XK

AAW3886 ck: 209 len: 32 ! Aaw3886 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20} VTK

AAW3852 ck: 5095 len: 41 ! Aaw3852 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
```

```
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
   (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK XK
   (K) {20}

AAW38887 ck: 2704 len: 33 1 Aaw38887 Delivery peptide used in peptide m
   (R,K) {20,20}
   (K) {20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
```

```
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKVTK
   (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KVTKK
   (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK VTK
   (K) {20}

AAW38853 ck: 8258 len: 42 1 Aaw38853 Delivery peptide used in peptide m
   (R,K) {20,20}
   (K) {20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
   (K) {20}
```

21: KKKK KKKKKKKKKKKKKKKKKKKKK KK

AAW38888 ck: 5274 len: 34 ! Aaw38888 Delivery peptide used in peptide m

(R,K){20,20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAW38854 ck: 1496 len: 43 ! Aaw38854 Delivery peptide used in peptide m

(R,K){20,20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAW38889 ck: 7919 len: 35 ! Aaw38889 Delivery peptide used in peptide m

(R,K){20,20}

(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

11: KKKK (K) {20}
12: KKKK (K) {20}
13: KKKK (K) {20} VTK

AAW38890 ck: 639 len: 36 ! Aaw38890 Delivery peptide used in peptide m

1

(R,K) {20,20}
1: KKKK (K) {20}
2: K KKKK (K) {20}
3: KK KKKK (K) {20}
4: KK KKKK (K) {20}
5: KKK KKKK (K) {20}
6: KKKK KKKK (K) {20}
7: KKKK KKKK (K) {20}
8: KKKK KKKK (K) {20}
9: KKKK KKKK (K) {20}
10: KKKK KKKK (K) {20}
11: KKKK KKKK (K) {20}
12: KKKK KKKK (K) {20}
13: KKKK KKKK (K) {20} KVT
14: KKKK KKKK (K) {20} VTK

AAW38891 ck: 3434 len: 37 ! Aaw38891 Delivery peptide used in peptide m

1

(R,K) {20,20}
1: KKKK (K) {20}
2: K KKKK (K) {20}
3: KK KKKK (K) {20}
4: KK KKKK (K) {20}
5: KKK KKKK (K) {20}

(K) {20}
6: KKKK KKKK (K) {20}
7: KKKK KKKK (K) {20}
8: KKKK KKKK (K) {20}
9: KKKK KKKK (K) {20}
10: KKKK KKKK (K) {20}
11: KKKK KKKK (K) {20}
12: KKKK KKKK (K) {20}
13: KKKK KKKK (K) {20} KVT
14: KKKK KKKK (K) {20} KVT
15: KKKK KKKK (K) {20} VTK

AAW38892 ck: 6304 len: 38 ! Aaw38892 Delivery peptide used in peptide m

1

(R,K) {20,20}
1: KKKK (K) {20}
2: K KKKK (K) {20}
3: KK KKKK (K) {20}
4: KK KKKK (K) {20}
5: KK KKKK (K) {20}
6: KKK KKKK (K) {20}
7: KKKK KKKK (K) {20}
8: KKKK KKKK (K) {20}
9: KKKK KKKK (K) {20}
10: KKKK KKKK (K) {20}
11: KKKK KKKK (K) {20}
12: KKKK KKKK (K) {20}
13: KKKK KKKK (K) {20} KVT

14: KKKK (K) {20} KKVTK
 15: KKKK (K) {20} KVTK
 16: KKKK (K) {20} VTK

AAW3893 ck: 9249 len: 39 ! Aaw3893 Delivery peptide used in peptide m

1

1: (R,K) {20,20} KKKK
 2: K (K) {20} KKKK
 3: KK (K) {20} KKKK
 4: KKK (K) {20} KKKK
 5: KKKK (K) {20} KKKK
 6: KKKK (K) {20} KKKK
 7: KKKK (K) {20} KKKK
 8: KKKK (K) {20} KKKK
 9: KKKK (K) {20} KKKK
 10: KKKK (K) {20} KKKK
 11: KKKK (K) {20} KKKK
 12: KKKK (K) {20} KKKK
 13: KKKK (K) {20} KKKK
 14: KKKK (K) {20} KKVTK
 15: KKKK (K) {20} KVTK
 16: KKKK (K) {20} KVTK
 17: KKKK (K) {20} VTK

AAW3894 ck: 2269 len: 40 ! Aaw3894 Delivery peptide used in peptide m

1

1: (R,K) {20,20} KKKK
 2: K (K) {20} KKKK
 3: KK (K) {20} KKKK
 4: KKK (K) {20} KKKK
 5: KKKK (K) {20} KKKK
 6: KKKK (K) {20} KKKK
 7: KKKK (K) {20} KKKK
 8: KKKK (K) {20} KKKK
 9: KKKK (K) {20} KKKK
 10: KKKK (K) {20} KKKK
 11: KKKK (K) {20} KKKK
 12: KKKK (K) {20} KKKK
 13: KKKK (K) {20} KKKK
 14: KKKK (K) {20} KKVTK
 15: KKKK (K) {20} KVTK
 16: KKKK (K) {20} KVTK
 17: KKKK (K) {20} VTK

(K) {20}

1

2: K KKKK (K) {20} KKKK
 3: KK KKKK (K) {20} KKKK
 4: KKK KKKK (K) {20} KKKK
 5: KKKK KKKK (K) {20} KKKK
 6: KKKK KKKK (K) {20} KKKK
 7: KKKK KKKK (K) {20} KKKK
 8: KKKK KKKK (K) {20} KKKK
 9: KKKK KKKK (K) {20} KKKK
 10: KKKK KKKK (K) {20} KKKK
 11: KKKK KKKK (K) {20} KKKK
 12: KKKK KKKK (K) {20} KKKK
 13: KKKK KKKK (K) {20} KKKK
 14: KKKK KKKK (K) {20} KKKK
 15: KKKK KKKK (K) {20} KKVTK
 16: KKKK KKKK (K) {20} KVTK
 17: KKKK KKKK (K) {20} VTK
 18: KKKK KKKK (K) {20} VTK

AAW3895 ck: 5364 len: 41 ! Aaw3895 Delivery peptide used in peptide m

(R,K) {20,20}

1: KKKK (K) {20} KKKK
 2: K KKKK (K) {20} KKKK
 3: KK KKKK (K) {20} KKKK
 4: KKK KKKK (K) {20} KKKK
 5: KKKK KKKK (K) {20} KKKK
 6: KKKK KKKK (K) {20} KKKK

(K) {20}

```

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}

AWW38896 ck: 8534 len: 42 ! AWW38896 Delivery peptide used in peptide n
(R,K) {20,20}

```

[illegible]

```

11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
19: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}

AAW38897 ck: 1779 len: 43 ! Aaw38897 Delivery peptide used in peptide me
1:
      (R, K) {20, 20}
      (K) {20}
      KKKKKKKKKKKKKKKKKKK KKKK
2: K KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
3: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
4: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
5: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
6: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
7: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
8: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
9: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
10: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
11: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
12: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
13: KK KKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}

```

```

14: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
15: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
16: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
17: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
18: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
19: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
20: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
21: KKKK KKKKKKKKKKKKKKKKKKKKK VTK

AAW38898 ck: 5099 len: 44 ! Aaw38898 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
4: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
5: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
6: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
7: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
8: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
9: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
10: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
11: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
12: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
13: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
14: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
15: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

```

```

16: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
17: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
18: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
19: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
20: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
21: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
22: KKKK KKKKKKKKKKKKKKKKKKKKK VTK

AAW38809 ck: 7658 len: 35 ! Aaw38809 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
4: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
5: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
6: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
7: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
8: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
9: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
10: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
11: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
12: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
13: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
14: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

AAW38810 ck: 370 len: 36 ! Aaw38810 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK

```

```
2: K (K) {20}
3: KK (K) {20}
4: KKK (K) {20}
5: KKKK (K) {20}
6: KKKKK (K) {20}
7: KKKKKK (K) {20}
8: KKKKKKK (K) {20}
9: KKKKKKKK (K) {20}
10: KKKKKKKKK (K) {20}
11: KKKKKKKKKK (K) {20}
12: KKKKKKKKKKK (K) {20}
13: KKKKKKKKKKKK (K) {20}
14: KKKKKKKKKKKKK (K) {20}
15: KKKKKKKKKKKKKK (K) {20}
```

AAW38811 ck: 3157 len: 37 ! Aaw38811 Delivery peptide used in peptide p

```
1: (R,K) {20,20}
2: K (K) {20}
3: KK (K) {20}
4: KKK (K) {20}
5: KKKK (K) {20}
6: KKKKK (K) {20}
7: KKKKKK (K) {20}
8: KKKKKKK (K) {20}
9: KKKKKKKK (K) {20}
```

```
10: KKKKK (K) {20}
11: KKKKKK (K) {20}
12: KKKKKKK (K) {20}
13: KKKKKKKK (K) {20}
14: KKKKKKKKK (K) {20}
15: KKKKKKKKKK (K) {20}
16: KKKKKKKKKKK (K) {20}
```

AAW38812 ck: 6019 len: 38 ! Aaw38812 Delivery peptide used in peptide m

```
1: (R,K) {20,20}
2: K (K) {20}
3: KK (K) {20}
4: KKK (K) {20}
5: KKKK (K) {20}
6: KKKKK (K) {20}
7: KKKKKK (K) {20}
8: KKKKKKK (K) {20}
9: KKKKKKKK (K) {20}
10: KKKKKKKKK (K) {20}
11: KKKKKKKKKK (K) {20}
12: KKKKKKKKKKK (K) {20}
13: KKKKKKKKKKKK (K) {20}
14: KKKKKKKKKKKKK (K) {20}
15: KKKKKKKKKKKKKK (K) {20}
16: KKKKKKKKKKKKKKK (K) {20}
```


17: KKKKK (K){20}
KKKKKKKKKKKKKKKKKKKK WK

AAW38813 ck: 8956 len: 39 1 Aaw38813 Delivery peptide used in peptide m

(R,K){20,20}

1: KKKKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKKKK KKKKK

3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKK KKKKKKKKKKKKKKKKKKKKK WK

AAW38814 ck: 1968 len: 40 1 Aaw38814 Delivery peptide used in peptide m

(R,K){20,20}

1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKK KKKKK

3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKK KKKKKKKKKKKKKKKKKKKKK WK

AAW38815 ck: 5055 len: 41 1 Aaw38815 Delivery peptide used in peptide m

(R,K){20,20}

1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKK KKKKK

3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK (K) {20}
 9: KKKKK (K) {20}
 10: KKKKK (K) {20}
 11: KKKKK (K) {20}
 12: KKKKK (K) {20}
 13: KKKKK (K) {20}
 14: KKKKK (K) {20}
 15: KKKKK (K) {20}
 16: KKKKK (K) {20}
 17: KKKKK (K) {20}
 18: KKKKK (K) {20}
 19: KKKKK (K) {20}
 20: KKKKK (K) {20} WK

1
 AAW38816 ck: 8217 len: 42 1 Aaw38816 Delivery peptide used in peptide m
 1: (R,K) {20,20}
 2: K (K) {20}
 3: KK (K) {20}
 4: KKK (K) {20}
 5: KKKK (K) {20}
 6: KKKKK (K) {20}
 7: KKKKK (K) {20}
 8: KKKKK (K) {20}
 9: KKKKK (K) {20}
 10: KKKKK (K) {20}

11: KKKKK (K) {20}
 12: KKKKK (K) {20}
 13: KKKKK (K) {20}
 14: KKKKK (K) {20}
 15: KKKKK (K) {20}
 16: KKKKK (K) {20}
 17: KKKKK (K) {20}
 18: KKKKK (K) {20}
 19: KKKKK (K) {20}
 20: KKKKK (K) {20}
 21: KKKKK (K) {20} WK

1
 AAW38817 ck: 1454 len: 43 1 Aaw38817 Delivery peptide used in peptide m
 1: (R,K) {20,20}
 2: K (K) {20}
 3: KK (K) {20}
 4: KKK (K) {20}
 5: KKKK (K) {20}
 6: KKKKK (K) {20}
 7: KKKKK (K) {20}
 8: KKKKK (K) {20}
 9: KKKKK (K) {20}
 10: KKKKK (K) {20}
 11: KKKKK (K) {20}
 12: KKKKK (K) {20}

13: KKKK (K) {20}
14: KKKK (K) {20}
15: KKKK (K) {20}
16: KKKK (K) {20}
17: KKKK (K) {20}
18: KKKK (K) {20}
19: KKKK (K) {20}
20: KKKK (K) {20}
21: KKKK (K) {20}
22: KKKK (K) {20}

1 AAW24865 ck: 2211 len: 40 ! Aaw24865 Bifunctional peptide I for binding
(R,K) {20,20}
21: YEDES KKKKKKKKKKKKKKKKKKK

1 AAW24450 ck: 8137 len: 45 ! Aaw24450 Nucleic acid (NA) binding peptide
(R,K) {20,20}
4: YKA KKKKKKKKKKKKKKKKKKK
5: YPAK KKKKKKKKKKKKKKKKKKK
6: YKAKK KKKKKKKKKKKKKKKKKKK
7: KAKKK KKKKKKKKKKKKKKKKKKK
8: AKKKK KKKKKKKKKKKKKKKKKKK
9: KKKKK KKKKKKKKKKKKKKKKKKK
10: KKKKK KKKKKKKKKKKKKKKKKKK
11: KKKKK KKKKKKKKKKKKKKKKKKK
12: KKKKK KKKKKKKKKKKKKKKKKKK
13: KKKKK KKKKKKKKKKKKKKKKKKK
(K) {20}

14: KKKK KKKKKKKKKKKKKKKKKKK
15: KKKK (K) {20}
16: KKKK (K) {20}
17: KKKK (K) {20}
18: KKKK (K) {20}
19: KKKK (K) {20}
20: KKKK (K) {20}
21: KKKK (K) {20}
22: KKKK (K) {20}
23: KKKK (K) {20}
24: KKKK (K) {20}

1 AAW21590 ck: 4875 len: 30 ! Aaw21590 Antibiotic potentiating peptide #2
(R,K) {20,20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
2: K KKKKKKKKKKKKKKKKKKK KKKKK
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
8: KKKKK KKKKKKKKKKKKKKKKKKK KKK
9: KKKKK KKKKKKKKKKKKKKKKKKK KK
10: KKKKK KKKKKKKKKKKKKKKKKKK K
11: KKKKK KKKKKKKKKKKKKKKKKKK
(K) {20}

1 AAW21591 ck: 5075 len: 434 ! Aaw21591 Antibiotic potentiating peptide #3
(R,K) {20,20}

1:	XXXXX (K) {20}
2:	K (K) {20}
3:	KK (K) {20}
4:	KKK (K) {20}
5:	XXXXX (K) {20}
6:	XXXXX (K) {20}
7:	XXXXX (K) {20}
8:	XXXXX (K) {20}
9:	XXXXX (K) {20}
10:	XXXXX (K) {20}
11:	XXXXX (K) {20}
12:	XXXXX (K) {20}
13:	XXXXX (K) {20}
14:	XXXXX (K) {20}
15:	XXXXX (K) {20}
16:	XXXXX (K) {20}
17:	XXXXX (K) {20}
18:	XXXXX (K) {20}
19:	XXXXX (K) {20}
20:	XXXXX (K) {20}
21:	XXXXX (K) {20}
22:	XXXXX (K) {20}
23:	XXXXX (K) {20}
24:	XXXXX (K) {20}

25:	XXXXX (K) {20}
26:	XXXXX (K) {20}
27:	XXXXX (K) {20}
28:	XXXXX (K) {20}
29:	XXXXX (K) {20}
30:	XXXXX (K) {20}
31:	XXXXX (K) {20}
32:	XXXXX (K) {20}
33:	XXXXX (K) {20}
34:	XXXXX (K) {20}
35:	XXXXX (K) {20}
36:	XXXXX (K) {20}
37:	XXXXX (K) {20}
38:	XXXXX (K) {20}
39:	XXXXX (K) {20}
40:	XXXXX (K) {20}
41:	XXXXX (K) {20}
42:	XXXXX (K) {20}
43:	XXXXX (K) {20}
44:	XXXXX (K) {20}
45:	XXXXX (K) {20}
46:	XXXXX (K) {20}
47:	XXXXX (K) {20}
48:	XXXXX (K) {20}
49:	XXXXX (K) {20}

50: XXXXX (K) {20} XXXXX
 51: XXXXX (K) {20} XXXXX
 52: XXXXX (K) {20} XXXXX
 53: XXXXX (K) {20} XXXXX
 54: XXXXX (K) {20} XXXXX
 55: XXXXX (K) {20} XXXXX
 56: XXXXX (K) {20} XXXXX
 57: XXXXX (K) {20} XXXXX
 58: XXXXX (K) {20} XXXXX
 59: XXXXX (K) {20} XXXXX
 60: XXXXX (K) {20} XXXXX
 61: XXXXX (K) {20} XXXXX
 62: XXXXX (K) {20} XXXXX
 63: XXXXX (K) {20} XXXXX
 64: XXXXX (K) {20} XXXXX
 65: XXXXX (K) {20} XXXXX
 66: XXXXX (K) {20} XXXXX
 67: XXXXX (K) {20} XXXXX
 68: XXXXX (K) {20} XXXXX
 69: XXXXX (K) {20} XXXXX
 70: XXXXX (K) {20} XXXXX
 71: XXXXX (K) {20} XXXXX
 72: XXXXX (K) {20} XXXXX
 73: XXXXX (K) {20} XXXXX

74: XXXXX (K) {20} XXXXX
 75: XXXXX (K) {20} XXXXX
 76: XXXXX (K) {20} XXXXX
 77: XXXXX (K) {20} XXXXX
 78: XXXXX (K) {20} XXXXX
 79: XXXXX (K) {20} XXXXX
 80: XXXXX (K) {20} XXXXX
 81: XXXXX (K) {20} XXXXX
 82: XXXXX (K) {20} XXXXX
 83: XXXXX (K) {20} XXXXX
 84: XXXXX (K) {20} XXXXX
 85: XXXXX (K) {20} XXXXX
 86: XXXXX (K) {20} XXXXX
 87: XXXXX (K) {20} XXXXX
 88: XXXXX (K) {20} XXXXX
 89: XXXXX (K) {20} XXXXX
 90: XXXXX (K) {20} XXXXX
 91: XXXXX (K) {20} XXXXX
 92: XXXXX (K) {20} XXXXX
 93: XXXXX (K) {20} XXXXX
 94: XXXXX (K) {20} XXXXX
 95: XXXXX (K) {20} XXXXX
 96: XXXXX (K) {20} XXXXX
 97: XXXXX (K) {20} XXXXX

147: KKKK (K) {20}
 148: KKKK (K) {20}
 149: KKKK (K) {20}
 150: KKKK (K) {20}
 151: KKKK (K) {20}
 152: KKKK (K) {20}
 153: KKKK (K) {20}
 154: KKKK (K) {20}
 155: KKKK (K) {20}
 156: KKKK (K) {20}
 157: KKKK (K) {20}
 158: KKKK (K) {20}
 159: KKKK (K) {20}
 160: KKKK (K) {20}
 161: KKKK (K) {20}
 162: KKKK (K) {20}
 163: KKKK (K) {20}
 164: KKKK (K) {20}
 165: KKKK (K) {20}
 166: KKKK (K) {20}
 167: KKKK (K) {20}
 168: KKKK (K) {20}
 169: KKKK (K) {20}
 170: KKKK (K) {20}

171: KKKK (K) {20}
 172: KKKK (K) {20}
 173: KKKK (K) {20}
 174: KKKK (K) {20}
 175: KKKK (K) {20}
 176: KKKK (K) {20}
 177: KKKK (K) {20}
 178: KKKK (K) {20}
 179: KKKK (K) {20}
 180: KKKK (K) {20}
 181: KKKK (K) {20}
 182: KKKK (K) {20}
 183: KKKK (K) {20}
 184: KKKK (K) {20}
 185: KKKK (K) {20}
 186: KKKK (K) {20}
 187: KKKK (K) {20}
 188: KKKK (K) {20}
 189: KKKK (K) {20}
 190: KKKK (K) {20}
 191: KKKK (K) {20}
 192: KKKK (K) {20}
 193: KKKK (K) {20}
 194: KKKK (K) {20}
 195: KKKK (K) {20}

196: KKKK (K) {20} KKKK
197: KKKK (K) {20} KKKK
198: KKKK (K) {20} KKKK
199: KKKK (K) {20} KKKK
200: KKKK (K) {20} KKKK
201: KKKK (K) {20} KKKK
202: KKKK (K) {20} KKKK
203: KKKK (K) {20} KKKK
204: KKKK (K) {20} KKKK
205: KKKK (K) {20} KKKK
206: KKKK (K) {20} KKKK
207: KKKK (K) {20} KKKK
208: KKKK (K) {20} KKKK
209: KKKK (K) {20} KKKK
210: KKKK (K) {20} KKKK
211: KKKK (K) {20} KKKK
212: KKKK (K) {20} KKKK
213: KKKK (K) {20} KKKK
214: KKKK (K) {20} KKKK
215: KKKK (K) {20} KKKK
216: KKKK (K) {20} KKKK
217: KKKK (K) {20} KKKK
218: KKKK (K) {20} KKKK
219: KKKK (K) {20} KKKK

220: KKKK (K) {20} KKKK
221: KKKK (K) {20} KKKK
222: KKKK (K) {20} KKKK
223: KKKK (K) {20} KKKK
224: KKKK (K) {20} KKKK
225: KKKK (K) {20} KKKK
226: KKKK (K) {20} KKKK
227: KKKK (K) {20} KKKK
228: KKKK (K) {20} KKKK
229: KKKK (K) {20} KKKK
230: KKKK (K) {20} KKKK
231: KKKK (K) {20} KKKK
232: KKKK (K) {20} KKKK
233: KKKK (K) {20} KKKK
234: KKKK (K) {20} KKKK
235: KKKK (K) {20} KKKK
236: KKKK (K) {20} KKKK
237: KKKK (K) {20} KKKK
238: KKKK (K) {20} KKKK
239: KKKK (K) {20} KKKK
240: KKKK (K) {20} KKKK
241: KKKK (K) {20} KKKK
242: KKKK (K) {20} KKKK
243: KKKK (K) {20} KKKK

244:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) [20]	
245:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
246:	XXXXX	(X) [20]	XXXXX
247:	XXXXX	(X) [20]	XXXXX
248:	XXXXX	(X) [20]	XXXXX
249:	XXXXX	(X) [20]	XXXXX
250:	XXXXX	(X) [20]	XXXXX
251:	XXXXX	(X) [20]	XXXXX
252:	XXXXX	(X) [20]	XXXXX
253:	XXXXX	(X) [20]	XXXXX
254:	XXXXX	(X) [20]	XXXXX
255:	XXXXX	(X) [20]	XXXXX
256:	XXXXX	(X) [20]	XXXXX
257:	XXXXX	(X) [20]	XXXXX
258:	XXXXX	(X) [20]	XXXXX
259:	XXXXX	(X) [20]	XXXXX
260:	XXXXX	(X) [20]	XXXXX
261:	XXXXX	(X) [20]	XXXXX
262:	XXXXX	(X) [20]	XXXXX
263:	XXXXX	(X) [20]	XXXXX
264:	XXXXX	(X) [20]	XXXXX
265:	XXXXX	(X) [20]	XXXXX
266:	XXXXX	(X) [20]	XXXXX
267:	XXXXX	(X) [20]	XXXXX
268:	XXXXX	(X) [20]	XXXXX

269:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
270:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
271:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
272:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
273:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
274:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
275:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
276:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
277:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
278:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
279:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
280:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
281:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
282:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
283:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
284:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
285:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
286:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
287:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
288:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
289:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
290:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
291:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	
292:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX
		(k) 120	

293: KKKK (K) {20} KKKK
294: KKKK (K) {20} KKKK
295: KKKK (K) {20} KKKK
296: KKKK (K) {20} KKKK
297: KKKK (K) {20} KKKK
298: KKKK (K) {20} KKKK
299: KKKK (K) {20} KKKK
300: KKKK (K) {20} KKKK
301: KKKK (K) {20} KKKK
302: KKKK (K) {20} KKKK
303: KKKK (K) {20} KKKK
304: KKKK (K) {20} KKKK
305: KKKK (K) {20} KKKK
306: KKKK (K) {20} KKKK
307: KKKK (K) {20} KKKK
308: KKKK (K) {20} KKKK
309: KKKK (K) {20} KKKK
310: KKKK (K) {20} KKKK
311: KKKK (K) {20} KKKK
312: KKKK (K) {20} KKKK
313: KKKK (K) {20} KKKK
314: KKKK (K) {20} KKKK
315: KKKK (K) {20} KKKK
316: KKKK (K) {20} KKKK

317: KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
318: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
319: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
320: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
321: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
322: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
323: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
324: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
325: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
326: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
327: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
328: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
329: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
330: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
331: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
332: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
333: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
334: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
335: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
336: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
337: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
338: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
339: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
340: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
341: KKKK (K) {20} KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK

342: KKKK (K) {20} KKKK
 343: KKKK (K) {20} KKKK
 344: KKKK (K) {20} KKKK
 345: KKKK (K) {20} KKKK
 346: KKKK (K) {20} KKKK
 347: KKKK (K) {20} KKKK
 348: KKKK (K) {20} KKKK
 349: KKKK (K) {20} KKKK
 350: KKKK (K) {20} KKKK
 351: KKKK (K) {20} KKKK
 352: KKKK (K) {20} KKKK
 353: KKKK (K) {20} KKKK
 354: KKKK (K) {20} KKKK
 355: KKKK (K) {20} KKKK
 356: KKKK (K) {20} KKKK
 357: KKKK (K) {20} KKKK
 358: KKKK (K) {20} KKKK
 359: KKKK (K) {20} KKKK
 360: KKKK (K) {20} KKKK
 361: KKKK (K) {20} KKKK
 362: KKKK (K) {20} KKKK
 363: KKKK (K) {20} KKKK
 364: KKKK (K) {20} KKKK
 365: KKKK (K) {20} KKKK

366: KKKK (K) {20} KKKK
 367: KKKK (K) {20} KKKK
 368: KKKK (K) {20} KKKK
 369: KKKK (K) {20} KKKK
 370: KKKK (K) {20} KKKK
 371: KKKK (K) {20} KKKK
 372: KKKK (K) {20} KKKK
 373: KKKK (K) {20} KKKK
 374: KKKK (K) {20} KKKK
 375: KKKK (K) {20} KKKK
 376: KKKK (K) {20} KKKK
 377: KKKK (K) {20} KKKK
 378: KKKK (K) {20} KKKK
 379: KKKK (K) {20} KKKK
 380: KKKK (K) {20} KKKK
 381: KKKK (K) {20} KKKK
 382: KKKK (K) {20} KKKK
 383: KKKK (K) {20} KKKK
 384: KKKK (K) {20} KKKK
 385: KKKK (K) {20} KKKK
 386: KKKK (K) {20} KKKK
 387: KKKK (K) {20} KKKK
 388: KKKK (K) {20} KKKK
 389: KKKK (K) {20} KKKK

390: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
391: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
392: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
393: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
394: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
395: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
396: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
397: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
398: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
399: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
400: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
401: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
402: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
403: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
404: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
405: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
406: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
407: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
408: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
409: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
410: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
411: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
412: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
413: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
414: KKKK KKKKKKKKKKKKKKKKKKKKK K

1
415: KKKK KKKKKKKKKKKKKKKKKKKKK
(K) {20}
AAW65939 ck: 1569 len: 40 ! Aaw65939 Polylysine peptide NBC32. 11/1998
(R,K) {20,20}
2: T KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
3: TK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
4: TK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
5: TKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
6: TKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
7: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
8: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
9: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
10: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
11: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
12: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
13: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
14: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
15: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
16: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
17: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
18: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
1
AAW48808 ck: 9658 len: 56 ! Aaw48808 Homo sapiens clone CG109_1 protein
(R,K) {20,20}
34: EEPRE KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
35: EPREK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
36: PREKK KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

30: KKKKK (K) {20} KKKKK
 31: KKKKK (K) {20} KKKKK
 32: KKKKK (K) {20} KKKKK
 33: KKKKK (K) {20} KKKKK
 34: KKKKK (K) {20} KKKKK
 35: KKKKK (K) {20} KKKKK
 36: KKKKK (K) {20} KKKKK
 37: KKKKK (K) {20} KKKKK
 38: KKKKK (K) {20} KKKKK
 39: KKKKK (K) {20} KKKKK
 40: KKKKK (K) {20} KKKKK

AAB53806 ck: 8373 len: 64 1 Aab53806 Human colon cancer antigen protein
 (R, K) {20, 20}

25: IDCKS KKKKK (K) {20} KKKKK
 26: DCDSK KKKKK (K) {20} KKKKK
 27: CDSKK KKKKK (K) {20} KKKKK
 28: DSKKK KKKKK (K) {20} KKKKK
 29: SKKKK KKKKK (K) {20} KKKKK
 30: KKKKK KKKKK (K) {20} KKKKK

AAB53977 ck: 7756 len: 75 1 Aab53977 Human colon cancer antigen protein
 (R, K) {20, 20}

24: QFCFL KKKKK (K) {20} KKKKK
 25: FCCLK KKKKK (K) {20} KKKKK
 26: CLKKK KKKKK (K) {20} KKKKK
 27: LKKKK KKKKK (K) {20} KKKKK

28: LKKKK KKKKK (K) {20} KKKKK
 29: KKKKK KKKKK (K) {20} KKKKK
 30: KKKKK KKKKK (K) {20} KKKKK
 31: KKKKK KKKKK (K) {20} KKKKK
 32: KKKKK KKKKK (K) {20} KKKKK
 33: KKKKK KKKKK (K) {20} KKKKK
 34: KKKKK KKKKK (K) {20} KKKKK
 35: KKKKK KKKKK (K) {20} KKKKK
 36: KKKKK KKKKK (K) {20} KKKKK
 37: KKKKK KKKKK (K) {20} KKKKK
 38: KKKKK KKKKK (K) {20} KKKKK
 39: KKKKK KKKKK (K) {20} KKKKK

40: KKKKK KKKKK (K) {20} KKKKK
 41: KKKKK KKKKK (K) {20} KKKKK
 42: KKKKK KKKKK (K) {20} KKKKK
 43: KKKKK KKKKK (K) {20} KKKKK
 44: KKKKK KKKKK (K) {20} KKKKK
 45: KKKKK KKKKK (K) {20} KKKKK
 46: KKKKK KKKKK (K) {20} KKKKK
 47: KKKKK KKKKK (K) {20} KKKKK
 48: KKKKK KKKKK (K) {20} KKKKK
 49: KKKKK KKKKK (K) {20} KKKKK

AAB53980 ck: 881 len: 45 1 Aab53980 Human colon cancer antigen protein
 (R, K) {20, 20}

8: TCSFQ KKKKK (K) {20} KKKKK

9: CSFQK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}
10: SFQKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
11: FQKKK KKKKKKKKKKKKKKKKKKKKK KXGGR
(K) {20}
12: QKKKK KKKKKKKKKKKKKKKKKKKKK XGGRF

AAB54314 ck: 6235 len: 55 ! Aab54314 Human pancreatic cancer antigen pr

(R,K) {20,20}

(K) {20}

24: FFXKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

25: FXXXX KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

26: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

27: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

28: XKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG

(K) {20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG

(K) {20}

30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGR

(K) {20}

31: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGRF

(K) {20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKK GGRF

AAB56121 ck: 5941 len: 125 ! Aab56121 Human secreted protein sequence en

(R,K) {20,20}

(K) {20}

87: ILKSQ KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

88: LKSQK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

89: KSQKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

90: SQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

91: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

92: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

93: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

94: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

95: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG

(K) {20}

100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG

(K) {20}

101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGR

(K) {20}

102: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGP

(K) {20}

103: KKKKK KKKKKKKKKKKKKKKKKKKKK GGP

AAB27956 ck: 6732 len: 139 ! Aab27956 Human secreted protein seq ID NO:

(R,K) {20,20}

(K) {20}

91: HFRPG KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

92: FRPGK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

93: RPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

94: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

95: GKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

102: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

103: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

104: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

105: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

```

106: KKKKK (K) {20}
107: KKKKK (K) {20}
108: KKKKK (K) {20}
109: KKKKK (K) {20}
110: KKKKK (K) {20}
111: KKKKK (K) {20}
112: KKKKK (K) {20}
113: KKKKK (K) {20}
114: KKKKK (K) {20}
115: KKKKK (K) {20}
116: KKKKK (K) {20}
117: KKKKK (K) {20}
118: KKKKK (K) {20}
119: KKKKK (K) {20}
120: KKKKK (K) {20}

AAB39140 ck: 9531 len: 66 1 Aab39140 Human secreted protein #48. 2/2001
(R,K) {20,20}
41: MMTVX KKKKKKKKKKKKKKKKKKKKK KKKKK
42: WTVXK KKKKKKKKKKKKKKKKKKKKK KKKKK
43: TVXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
44: VXXKK KKKKKKKKKKKKKKKKKKKKK KKK
45: XXXKK KKKKKKKKKKKKKKKKKKKKK KK
46: KKKKK KKKKKKKKKKKKKKKKKKKKK K
47: KKKKK KKKKKKKKKKKKKKKKKKKKK

AAB41457 ck: 4536 len: 168 1 Aab41457 Human ORF ORF1221 polypeptide seq

```

```

144: KQOKP (R,K) {20,20}
145: QOKPK (K) {20}
146: QPKPK (K) {20}
147: KPKKK (K) {20}
148: PKKKK (K) {20}
149: KKKKK (K) {20}

AAB42786 ck: 4959 len: 102 1 Aab42786 Human ORF ORF2550 polypeptide seq
(R,K) {20,20}
81: SDVLQ KKKKKKKKKKKKKKKKKKKKK KK
82: DVLQK KKKKKKKKKKKKKKKKKKKKK K
83: VLQKK KKKKKKKKKKKKKKKKKKKKK

AAB42889 ck: 6432 len: 62 1 Aab42889 Human ORF ORF2653 polypeptide seq
(R,K) {20,20}
40: SHRCL KKKKKKKKKKKKKKKKKKKKK KKK
41: HRCLK KKKKKKKKKKKKKKKKKKKKK KK
42: RCLKK KKKKKKKKKKKKKKKKKKKKK K
43: CLKKK KKKKKKKKKKKKKKKKKKKKK

AAB43641 ck: 5216 len: 133 1 Aab43641 Human cancer associated protein seq
(R,K) {20,20}
114: GHEQS KKKKKKKKKKKKKKKKKKKKK

AAB43835 ck: 4025 len: 223 1 Aab43835 Human cancer associated protein seq
(R,K) {20,20}
196: NILFW KKKKKKKKKKKKKKKKKKKKK XXGGA

AAB44188 ck: 4991 len: 43 1 Aab44188 Human cancer associated protein seq
(R,K) {20,20}
2: E KKKKKKKKKKKKKKKKKKKKK KKKKK
3: EK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}

```


4: /EKK KKKKKKKKKKKKKKKKKKKKK KKKGR
(K) {20}
5: EKKK KKKKKKKKKKKKKKKKKKKKK KKKRP
(K) {20}
6: EKKK KKKKKKKKKKKKKKKKKKKKK KGRPF
(K) {20}
7: KKKK KKKKKKKKKKKKKKKKKKKKK GRPF
(K) {20}

AAB4380 ck: 8490 len: 42 ! Aab4380 Human secreted protein encoded by
(R,K) {20,20}
(K) {20}

22: IWKI KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

23: WKIK KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAB4331 ck: 8131 len: 66 ! Aab4331 Human secreted protein sequence en
(R,K) {20,20}
(K) {20}

40: IHHP KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

41: HHKP KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

42: HKPK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

43: KPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

44: PKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}

45: KKKK KKKKKKKKKKKKKKKKKKKKK KK
(K) {20}

46: KKKK KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

47: KKKK KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAB3585 ck: 9893 len: 36 ! Aab3585 AbK21 linker peptide. 1/2001
(R,K) {20,20}
(K) {20}

15: PGSG KKKKKKKKKKKKKKKKKKKKK KG
(K) {20}

16: GSGK KKKKKKKKKKKKKKKKKKKKK G
(K) {20}

AAB3586 ck: 58 len: 36 ! Aab3586 GeK21 linker peptide. 1/2001
(R,K) {20,20}
(K) {20}

15: SGSG KKKKKKKKKKKKKKKKKKKKK KG
(K) {20}

16: GSGK KKKKKKKKKKKKKKKKKKKKK G
(K) {20}

AAB3591 ck: 7684 len: 630 ! Aab3591 Modified fibre protein encoded in
(R,K) {20,20}

596: PGSG KKKKKKKKKKKKKKKKKKKKK KGSYS
(K) {20}

597: GSGK KKKKKKKKKKKKKKKKKKKKK GSYS
(K) {20}

AAB3592 ck: 7647 len: 630 ! Aab3592 Modified fibre protein encoded in
(R,K) {20,20}
(K) {20}

596: GSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
(K) {20}

597: GSGK KKKKKKKKKKKKKKKKKKKKK GSYS
(K) {20}

AAB3593 ck: 2249 len: 640 ! Aab3593 Modified fibre protein encoded in
(R,K) {20,20}
(K) {20}

596: PGSG KKKKKKKKKKKKKKKKKKKKK KGSAB
(K) {20}

597: GSGK KKKKKKKKKKKKKKKKKKKKK GSAEK
(K) {20}

AAB3594 ck: 2612 len: 640 ! Aab3594 Modified fibre protein encoded in
(R,K) {20,20}
(K) {20}

596: SGSG KKKKKKKKKKKKKKKKKKKKK KGSAB
(K) {20}

597: GSGK KKKKKKKKKKKKKKKKKKKKK GSAEK
(K) {20}

AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen pK. 11/2001
(R,K) {20,20}
(K) {20}

2: C KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAB13783 ck: 4553 len: 45 ! Aab13783 Soluble tandem pBA/ pK peptide con
(R,K) {20,20}
(K) {20}

26: AAAA KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAB13784 ck: 4126 len: 44 ! Aab13784 Soluble tandem HA/ pK peptide conj
(R,K) {20,20}
(K) {20}

25: DMYG KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAG00834 ck: 6330 len: 103 ! Aag00834 Human secreted protein, SEQ ID NO:
(R,K) {20,20}
(K) {20}

83: RKRQ KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

84: KRQK KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAV98493 ck: 8137 len: 45 ! Aav98493 Peptide #5 used in nucleic acid tr
(R,K) {20,20}
(K) {20}

4: YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

```

5: YRAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
6: YRAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
7: KAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
8: AKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
9: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
10: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
11: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
12: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
13: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
14: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
15: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
16: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
17: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
18: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
19: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
20: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
21: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
22: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
23: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
24: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAV98495 ck: 4361 len: 59 1 Aay98495 Nuclear ligand used in nucleic ac
(R,K) {20,20}
18: APYA KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
19: PYAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
20: YRAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

```

[illegible]

6: KKKK (K) {20} KKKK
 7: KKKK (K) {20} KKKK
 8: KKKK (K) {20} KKKK
 9: KKKK (K) {20} KKKK
 10: KKKK (K) {20} KKKK
 11: KKKK (K) {20} KKKK
 12: KKKK (K) {20} KKKK
 13: KKKK (K) {20} KKKK
 14: KKKK (K) {20} KKKK
 15: KKKK (K) {20} KKKK
 16: KKKK (K) {20} KKKK
 17: KKKK (K) {20} KKKK
 18: KKKK (K) {20} KKKK
 19: KKKK (K) {20} KKKK
 20: KKKK (K) {20} KKKK
 21: KKKK (K) {20} KKKK
 22: KKKK (K) {20} KKKK
 23: KKKK (K) {20} KKKK
 24: KKKK (K) {20} KKKK
 25: KKKK (K) {20} KKKK
 26: KKKK (K) {20} KKKK
 27: KKKK (K) {20} KKKK
 28: KKKK (K) {20} KKKK
 29: KKKK (K) {20} KKKK

30: KKKK (K) {20} KKKK
 31: KKKK (K) {20} KKKK
 32: KKKK (K) {20} KKKK
 33: KKKK (K) {20} KKKK
 34: KKKK (K) {20} KKKK
 35: KKKK (K) {20} KKKK
 36: KKKK (K) {20} KKKK
 37: KKKK (K) {20} KKKK
 38: KKKK (K) {20} KKKK
 39: KKKK (K) {20} KKKK
 40: KKKK (K) {20} KKKK
 41: KKKK (K) {20} KKKK
 42: KKKK (K) {20} KKKK
 43: KKKK (K) {20} KKKK
 44: KKKK (K) {20} KKKK
 45: KKKK (K) {20} KKKK
 46: KKKK (K) {20} KKKK
 47: KKKK (K) {20} KKKK
 48: KKKK (K) {20} KKKK
 49: KKKK (K) {20} KKKK
 50: KKKK (K) {20} KKKK
 51: KKKK (K) {20} KKKK
 52: KKKK (K) {20} KKKK
 53: KKKK (K) {20} KKKK

```

54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
72: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
73: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
74: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
75: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
76: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
77: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
78: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
      (K) {20}

```

```

1
AY81805 ck: 1885 len: 351 1 AY81805 Murine mahogany protein sequence at
      (R,K) {20,20}
      (R,K) {20}
3: EF KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
4: EFR KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
5: EFRK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
6: EFRKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
7: FRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
8: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}

```

23: KKKK (K) {20}
 24: KKKK (K) {20}
 25: KKKK (K) {20}
 26: KKKK (K) {20}
 27: KKKK (K) {20}

AAy56902 ck: 4875 len: 30 ! Aay56902 (Lys)30 protein sequence. 4/2000

1

1: KKKK (R, K) {20, 20}
 2: K (K) {20}
 3: KK (K) {20}
 4: KKK (K) {20}
 5: KKKK (K) {20}
 6: KKKK (K) {20}
 7: KKKK (K) {20}
 8: KKKK (K) {20}
 9: KKKK (K) {20}
 10: KKKK (K) {20}
 11: KKKK (K) {20}

AAy56903 ck: 5075 len: 434 ! Aay56903 (Lys)434 protein sequence. 4/2000

1

1: KKKK (R, K) {20, 20}
 2: K (K) {20}
 3: KK (K) {20}
 4: KKK (K) {20}
 5: KKKK (K) {20}

6: KKKK (K) {20}
 7: KKKK (K) {20}
 8: KKKK (K) {20}
 9: KKKK (K) {20}
 10: KKKK (K) {20}
 11: KKKK (K) {20}
 12: KKKK (K) {20}
 13: KKKK (K) {20}
 14: KKKK (K) {20}
 15: KKKK (K) {20}
 16: KKKK (K) {20}
 17: KKKK (K) {20}
 18: KKKK (K) {20}
 19: KKKK (K) {20}
 20: KKKK (K) {20}
 21: KKKK (K) {20}
 22: KKKK (K) {20}
 23: KKKK (K) {20}
 24: KKKK (K) {20}
 25: KKKK (K) {20}
 26: KKKK (K) {20}
 27: KKKK (K) {20}
 28: KKKK (K) {20}
 29: KKKK (K) {20}
 30: KKKK (K) {20}

31: KKKK (K) {20} KKKK
32: KKKK (K) {20} KKKK
33: KKKK (K) {20} KKKK
34: KKKK (K) {20} KKKK
35: KKKK (K) {20} KKKK
36: KKKK (K) {20} KKKK
37: KKKK (K) {20} KKKK
38: KKKK (K) {20} KKKK
39: KKKK (K) {20} KKKK
40: KKKK (K) {20} KKKK
41: KKKK (K) {20} KKKK
42: KKKK (K) {20} KKKK
43: KKKK (K) {20} KKKK
44: KKKK (K) {20} KKKK
45: KKKK (K) {20} KKKK
46: KKKK (K) {20} KKKK
47: KKKK (K) {20} KKKK
48: KKKK (K) {20} KKKK
49: KKKK (K) {20} KKKK
50: KKKK (K) {20} KKKK
51: KKKK (K) {20} KKKK
52: KKKK (K) {20} KKKK
53: KKKK (K) {20} KKKK
54: KKKK (K) {20} KKKK

55: KKKK (K) {20} KKKK
56: KKKK (K) {20} KKKK
57: KKKK (K) {20} KKKK
58: KKKK (K) {20} KKKK
59: KKKK (K) {20} KKKK
60: KKKK (K) {20} KKKK
61: KKKK (K) {20} KKKK
62: KKKK (K) {20} KKKK
63: KKKK (K) {20} KKKK
64: KKKK (K) {20} KKKK
65: KKKK (K) {20} KKKK
66: KKKK (K) {20} KKKK
67: KKKK (K) {20} KKKK
68: KKKK (K) {20} KKKK
69: KKKK (K) {20} KKKK
70: KKKK (K) {20} KKKK
71: KKKK (K) {20} KKKK
72: KKKK (K) {20} KKKK
73: KKKK (K) {20} KKKK
74: KKKK (K) {20} KKKK
75: KKKK (K) {20} KKKK
76: KKKK (K) {20} KKKK
77: KKKK (K) {20} KKKK
78: KKKK (K) {20} KKKK

79: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 80: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 81: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 82: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 83: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 84: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 85: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 86: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 87: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 88: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 89: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 90: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 91: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 92: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 93: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 94: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 95: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 96: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 97: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 98: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 99: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 100: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 101: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 102: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 103: KKKK KKKKKKKKKKKKKKKKKKK KKKK

104: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 105: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 106: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 107: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 108: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 109: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 110: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 111: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 112: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 113: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 114: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 115: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 116: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 117: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 118: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 119: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 120: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 121: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 122: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 123: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 124: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 125: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 126: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 127: KKKK KKKKKKKKKKKKKKKKKKK KKKK

128: KKKK (K) {20}
 129: KKKK (K) {20}
 130: KKKK (K) {20}
 131: KKKK (K) {20}
 132: KKKK (K) {20}
 133: KKKK (K) {20}
 134: KKKK (K) {20}
 135: KKKK (K) {20}
 136: KKKK (K) {20}
 137: KKKK (K) {20}
 138: KKKK (K) {20}
 139: KKKK (K) {20}
 140: KKKK (K) {20}
 141: KKKK (K) {20}
 142: KKKK (K) {20}
 143: KKKK (K) {20}
 144: KKKK (K) {20}
 145: KKKK (K) {20}
 146: KKKK (K) {20}
 147: KKKK (K) {20}
 148: KKKK (K) {20}
 149: KKKK (K) {20}
 150: KKKK (K) {20}
 151: KKKK (K) {20}

152: KKKK (K) {20}
 153: KKKK (K) {20}
 154: KKKK (K) {20}
 155: KKKK (K) {20}
 156: KKKK (K) {20}
 157: KKKK (K) {20}
 158: KKKK (K) {20}
 159: KKKK (K) {20}
 160: KKKK (K) {20}
 161: KKKK (K) {20}
 162: KKKK (K) {20}
 163: KKKK (K) {20}
 164: KKKK (K) {20}
 165: KKKK (K) {20}
 166: KKKK (K) {20}
 167: KKKK (K) {20}
 168: KKKK (K) {20}
 169: KKKK (K) {20}
 170: KKKK (K) {20}
 171: KKKK (K) {20}
 172: KKKK (K) {20}
 173: KKKK (K) {20}
 174: KKKK (K) {20}
 175: KKKK (K) {20}
 176: KKKK (K) {20}

177: KKKK (K) {20} KKKK
178: KKKK (K) {20} KKKK
179: KKKK (K) {20} KKKK
180: KKKK (K) {20} KKKK
181: KKKK (K) {20} KKKK
182: KKKK (K) {20} KKKK
183: KKKK (K) {20} KKKK
184: KKKK (K) {20} KKKK
185: KKKK (K) {20} KKKK
186: KKKK (K) {20} KKKK
187: KKKK (K) {20} KKKK
188: KKKK (K) {20} KKKK
189: KKKK (K) {20} KKKK
190: KKKK (K) {20} KKKK
191: KKKK (K) {20} KKKK
192: KKKK (K) {20} KKKK
193: KKKK (K) {20} KKKK
194: KKKK (K) {20} KKKK
195: KKKK (K) {20} KKKK
196: KKKK (K) {20} KKKK
197: KKKK (K) {20} KKKK
198: KKKK (K) {20} KKKK
199: KKKK (K) {20} KKKK
200: KKKK (K) {20} KKKK

201: KKKK (K) {20} KKKK
202: KKKK (K) {20} KKKK
203: KKKK (K) {20} KKKK
204: KKKK (K) {20} KKKK
205: KKKK (K) {20} KKKK
206: KKKK (K) {20} KKKK
207: KKKK (K) {20} KKKK
208: KKKK (K) {20} KKKK
209: KKKK (K) {20} KKKK
210: KKKK (K) {20} KKKK
211: KKKK (K) {20} KKKK
212: KKKK (K) {20} KKKK
213: KKKK (K) {20} KKKK
214: KKKK (K) {20} KKKK
215: KKKK (K) {20} KKKK
216: KKKK (K) {20} KKKK
217: KKKK (K) {20} KKKK
218: KKKK (K) {20} KKKK
219: KKKK (K) {20} KKKK
220: KKKK (K) {20} KKKK
221: KKKK (K) {20} KKKK
222: KKKK (K) {20} KKKK
223: KKKK (K) {20} KKKK
224: KKKK (K) {20} KKKK

225:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
226:	XXXXX	(X) [20]	XXXXX
227:	XXXXX	(X) [20]	XXXXX
228:	XXXXX	(X) [20]	XXXXX
229:	XXXXX	(X) [20]	XXXXX
230:	XXXXX	(X) [20]	XXXXX
231:	XXXXX	(X) [20]	XXXXX
232:	XXXXX	(X) [20]	XXXXX
233:	XXXXX	(X) [20]	XXXXX
234:	XXXXX	(X) [20]	XXXXX
235:	XXXXX	(X) [20]	XXXXX
236:	XXXXX	(X) [20]	XXXXX
237:	XXXXX	(X) [20]	XXXXX
238:	XXXXX	(X) [20]	XXXXX
239:	XXXXX	(X) [20]	XXXXX
240:	XXXXX	(X) [20]	XXXXX
241:	XXXXX	(X) [20]	XXXXX
242:	XXXXX	(X) [20]	XXXXX
243:	XXXXX	(X) [20]	XXXXX
244:	XXXXX	(X) [20]	XXXXX
245:	XXXXX	(X) [20]	XXXXX
246:	XXXXX	(X) [20]	XXXXX
247:	XXXXX	(X) [20]	XXXXX
248:	XXXXX	(X) [20]	XXXXX
249:	XXXXX	(X) [20]	XXXXX

250:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXX
		(k) {20}	
251:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
252:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
253:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
254:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
255:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
256:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
257:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
258:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
259:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
260:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
261:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
262:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
263:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
264:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
265:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
266:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
267:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
268:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
269:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
270:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
271:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
272:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	
273:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXX
		(k) {20}	

274:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
275:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
276:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
277:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
278:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
279:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
280:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
281:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
282:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
283:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
284:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
285:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
286:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
287:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
288:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
289:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
290:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
291:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
292:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
293:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
294:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
295:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
296:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
297:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	

298:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
299:	XXXXX	(k) {20}	XXXXX
300:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
301:	XXXXX	(k) {20}	XXXXX
302:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
303:	XXXXX	(k) {20}	XXXXX
304:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
305:	XXXXX	(k) {20}	XXXXX
306:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
307:	XXXXX	(k) {20}	XXXXX
308:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
309:	XXXXX	(k) {20}	XXXXX
310:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
311:	XXXXX	(k) {20}	XXXXX
312:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
313:	XXXXX	(k) {20}	XXXXX
314:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
315:	XXXXX	(k) {20}	XXXXX
316:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
317:	XXXXX	(k) {20}	XXXXX
318:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
319:	XXXXX	(k) {20}	XXXXX
320:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
321:	XXXXX	(k) {20}	XXXXX
322:	XXXXX	(k) {20}	XXXXX

323: KKKKK (K) {20} KKKKK
 324: KKKKK (K) {20} KKKKK
 325: KKKKK (K) {20} KKKKK
 326: KKKKK (K) {20} KKKKK
 327: KKKKK (K) {20} KKKKK
 328: KKKKK (K) {20} KKKKK
 329: KKKKK (K) {20} KKKKK
 330: KKKKK (K) {20} KKKKK
 331: KKKKK (K) {20} KKKKK
 332: KKKKK (K) {20} KKKKK
 333: KKKKK (K) {20} KKKKK
 334: KKKKK (K) {20} KKKKK
 335: KKKKK (K) {20} KKKKK
 336: KKKKK (K) {20} KKKKK
 337: KKKKK (K) {20} KKKKK
 338: KKKKK (K) {20} KKKKK
 339: KKKKK (K) {20} KKKKK
 340: KKKKK (K) {20} KKKKK
 341: KKKKK (K) {20} KKKKK
 342: KKKKK (K) {20} KKKKK
 343: KKKKK (K) {20} KKKKK
 344: KKKKK (K) {20} KKKKK
 345: KKKKK (K) {20} KKKKK
 346: KKKKK (K) {20} KKKKK

347: KKKKK (K) {20} KKKKK
 348: KKKKK (K) {20} KKKKK
 349: KKKKK (K) {20} KKKKK
 350: KKKKK (K) {20} KKKKK
 351: KKKKK (K) {20} KKKKK
 352: KKKKK (K) {20} KKKKK
 353: KKKKK (K) {20} KKKKK
 354: KKKKK (K) {20} KKKKK
 355: KKKKK (K) {20} KKKKK
 356: KKKKK (K) {20} KKKKK
 357: KKKKK (K) {20} KKKKK
 358: KKKKK (K) {20} KKKKK
 359: KKKKK (K) {20} KKKKK
 360: KKKKK (K) {20} KKKKK
 361: KKKKK (K) {20} KKKKK
 362: KKKKK (K) {20} KKKKK
 363: KKKKK (K) {20} KKKKK
 364: KKKKK (K) {20} KKKKK
 365: KKKKK (K) {20} KKKKK
 366: KKKKK (K) {20} KKKKK
 367: KKKKK (K) {20} KKKKK
 368: KKKKK (K) {20} KKKKK
 369: KKKKK (K) {20} KKKKK
 370: KKKKK (K) {20} KKKKK

371: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
372: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
373: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
374: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
375: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
376: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
377: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
378: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
379: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
380: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
381: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
382: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
383: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
384: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
385: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
386: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
387: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
388: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
389: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
390: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
391: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
392: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
393: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
394: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
395: KKKK KKKKKKKKKKKKKKKKKKK KKKK

. 1

396: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
397: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
398: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
399: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
400: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
401: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
402: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
403: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
404: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
405: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
406: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
407: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
408: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
409: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
410: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
411: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
412: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
413: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
414: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
415: KKKK KKKKKKKKKKKKKKKKKKK KKKK

AAV86248 ck: 8783 len: 128 ! Aay86248 Human secreted protein HCHPP68, SR

(R,K) {20,20}

(K) {20}

84: ANPP KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

85: NPPK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

86: PPPK KKKKKKKKKKKKKKKKKKK KKKK

```

87: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
88: PKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
91: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
93: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
94: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
95: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
104: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
105: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
106: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}

AAV59038 ck: 8137 len: 45 ! Aay59038 Peptide used in the construction c
(R,K) {20,20}
(K) {20}
4: YKA KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
5: YKAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}
6: YKAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
      (K) {20}

```

```

7: KAKKK (K) {20}
8: AKKKK (K) {20}
9: KKKKK (K) {20}
10: KKKKK (K) {20}
11: KKKKK (K) {20}
12: KKKKK (K) {20}
13: KKKKK (K) {20}
14: KKKKK (K) {20}
15: KKKKK (K) {20}
16: KKKKK (K) {20}
17: KKKKK (K) {20}
18: KKKKK (K) {20}
19: KKKKK (K) {20}
20: KKKKK (K) {20}
21: KKKKK (K) {20}
22: KKKKK (K) {20}
23: KKKKK (K) {20}
24: KKKKK (K) {20}

AAVS9040 ck: 4361 len: 59 1 AayS9040 Nuclear ligand for transporting nu
(R,K) {20,20}
18: APYKA KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: PYKAK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: YKAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: KAKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: AKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

```

23: KKKK (K) {20} KKKK
24: KKKK (K) {20} KKKK
25: KKKK (K) {20} KKKK
26: KKKK (K) {20} KKKK
27: KKKK (K) {20} KKKK
28: KKKK (K) {20} KKKK
29: KKKK (K) {20} KKKK
30: KKKK (K) {20} KKKK
31: KKKK (K) {20} KKKK
32: KKKK (K) {20} KKKK
33: KKKK (K) {20} KKKK
34: KKKK (K) {20} KKKK
35: KKKK (K) {20} KKKK
36: KKKK (K) {20} KKKK
37: KKKK (K) {20} KKKK
38: KKKK (K) {20} KKKK

AA59042 ck: 4925 len: 100 ! Aay59042 Amino acid polymer seq ID NO: 62 c

1
1: (R,K) {20,20} KKKK
2: K (K) {20} KKKK
3: KK (K) {20} KKKK
4: KK (K) {20} KKKK
5: KKK (K) {20} KKKK
6: KKKK (K) {20} KKKK
7: KKKK (K) {20} KKKK

8: KKKK (K) {20} KKKK
9: KKKK (K) {20} KKKK
10: KKKK (K) {20} KKKK
11: KKKK (K) {20} KKKK
12: KKKK (K) {20} KKKK
13: KKKK (K) {20} KKKK
14: KKKK (K) {20} KKKK
15: KKKK (K) {20} KKKK
16: KKKK (K) {20} KKKK
17: KKKK (K) {20} KKKK
18: KKKK (K) {20} KKKK
19: KKKK (K) {20} KKKK
20: KKKK (K) {20} KKKK
21: KKKK (K) {20} KKKK
22: KKKK (K) {20} KKKK
23: KKKK (K) {20} KKKK
24: KKKK (K) {20} KKKK
25: KKKK (K) {20} KKKK
26: KKKK (K) {20} KKKK
27: KKKK (K) {20} KKKK
28: KKKK (K) {20} KKKK
29: KKKK (K) {20} KKKK
30: KKKK (K) {20} KKKK
31: KKKK (K) {20} KKKK

```
32: KKKK (K) {20} KKKK
33: KKKK (K) {20} KKKK
34: KKKK (K) {20} KKKK
35: KKKK (K) {20} KKKK
36: KKKK (K) {20} KKKK
37: KKKK (K) {20} KKKK
38: KKKK (K) {20} KKKK
39: KKKK (K) {20} KKKK
40: KKKK (K) {20} KKKK
41: KKKK (K) {20} KKKK
42: KKKK (K) {20} KKKK
43: KKKK (K) {20} KKKK
44: KKKK (K) {20} KKKK
45: KKKK (K) {20} KKKK
46: KKKK (K) {20} KKKK
47: KKKK (K) {20} KKKK
48: KKKK (K) {20} KKKK
49: KKKK (K) {20} KKKK
50: KKKK (K) {20} KKKK
51: KKKK (K) {20} KKKK
52: KKKK (K) {20} KKKK
53: KKKK (K) {20} KKKK
54: KKKK (K) {20} KKKK
55: KKKK (K) {20} KKKK

56: KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK KKKK
57: KKKK (K) {20} KKKK
58: KKKK (K) {20} KKKK
59: KKKK (K) {20} KKKK
60: KKKK (K) {20} KKKK
61: KKKK (K) {20} KKKK
62: KKKK (K) {20} KKKK
63: KKKK (K) {20} KKKK
64: KKKK (K) {20} KKKK
65: KKKK (K) {20} KKKK
66: KKKK (K) {20} KKKK
67: KKKK (K) {20} KKKK
68: KKKK (K) {20} KKKK
69: KKKK (K) {20} KKKK
70: KKKK (K) {20} KKKK
71: KKKK (K) {20} KKKK
72: KKKK (K) {20} KKKK
73: KKKK (K) {20} KKKK
74: KKKK (K) {20} KKKK
75: KKKK (K) {20} KKKK
76: KKKK (K) {20} KKKK
77: KKKK (K) {20} KKKK
78: KKKK (K) {20} KKKK
79: KKKK (K) {20} KKKK
80: KKKK (K) {20} K
```


19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABP66883 ck: 3983 len: 292 ! Abp66883 Human polypeptide SEQ ID NO 604. 1
(R,K) {20,20}
(R,K) {20}
273: QVPAP KKKKKKKKKKKKKKKKKKK

ABP67072 ck: 7611 len: 315 ! Abp67072 Human polypeptide SEQ ID NO 793. 1
(R,K) {20,20}
(R,K) {20}
273: QVPAP KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
274: VFPAP KKKKKKKKKKKKKKKKKKK KKKKK

ABG92583 ck: 7907 len: 39 ! Abg92583 Human DNA-binding protein #9. 11/2
(R,K) {20,20}
(K) {20}
9: YFEDL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
10: FEDLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: EDLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
12: DLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
13: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

ABG92588 ck: 9194 len: 87 ! Abg92588 Human DNA-binding protein #14. 11/
(R,K) {20,20}
(K) {20}
52: KILLL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
53: IILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
54: ILLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

56: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG92589 ck: 8659 len: 104 ! Abg92589 Human DNA-binding protein #15. 11/
(R,K) {20,20}
(K) {20}
75: PLGGQ KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
76: LGGQK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
77: GGQKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
78: GQKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
79: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

ABG92592 ck: 9398 len: 48 ! Abg92592 Human DNA-binding protein #18. 11/
(R,K) {20,20}
(K) {20}
2: Q KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
3: QK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
4: QKK KKKKKKKKKKKKKKKKKKK KKKKK

ABG92598 ck: 8278 len: 53 ! Abg92598 Human DNA-binding protein #24. 11/
(R,K) {20,20}
(K) {20}
30: NCGIL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
31: CGILK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
32: GILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
33: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK

ABG92599 ck: 444 len: 66 ! Abg92599 Human DNA-binding protein #25. 11/
(R,K) {20,20}
(K) {20}
35: SMTFS KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: MTFSK KKKKKKKKKKKKKKKKKKK KKKKK

ABG92600 ck: 5503 len: 50 ! Abg92600 Human DNA-binding protein #26. 11/
(R,K) {20,20}

30: IICLL (K) {20} K
31: ICLLK (K) {20} K

ABG92605 ck: 5691 len: 108 ! Abg92605 Human DNA-binding protein #31. 11/

(R,K) {20,20}

78: VRPCL (K) {20} K

79: RPCLK (K) {20} K

80: PCLLK (K) {20} K

81: CLKKK (K) {20} K

82: LKKKK (K) {20} K

83: KKKKK (K) {20} K

84: KKKKK (K) {20} K

85: KKKKK (K) {20} K

86: KKKKK (K) {20} K

87: KKKKK (K) {20} K

88: KKKKK (K) {20} K

89: KKKKK (K) {20} K

ABG92613 ck: 6029 len: 63 ! Abg92613 Human DNA-binding protein #39. 11/

(R,K) {20,20}

40: KLTLL (K) {20} K

ABG92621 ck: 7170 len: 63 ! Abg92621 Human DNA-binding protein #47. 11/

(R,K) {20,20}

37: TPERS (K) {20} K

38: PSRAK (K) {20} K

39: SRAKK (K) {20} K

40: RAKKK (K) {20} K

41: AKKKK (K) {20} K

42: KKKKK (K) {20} K

43: KKKKK (K) {20} K

44: KKKKK (K) {20} K

ABG92625 ck: 6110 len: 61 ! Abg92625 Human DNA-binding protein #51. 11/

(R,K) {20,20}

28: RPTRP (K) {20} K

29: PTRPK (K) {20} K

30: TRPKK (K) {20} K

31: RPKKK (K) {20} K

32: PKKKK (K) {20} K

33: KKKKK (K) {20} K

34: KKKKK (K) {20} K

35: KKKKK (K) {20} K

36: KKKKK (K) {20} K

37: KKKKK (K) {20} K

ABG92626 ck: 5764 len: 74 ! Abg92626 Human DNA-binding protein #52. 11/

(R,K) {20,20}

40: EFLSA (K) {20} K

41: FLSAK (K) {20} K

42: LSAAK (K) {20} K

43: SAKKK (K) {20} K

44: AKKKK (K) {20} K

45: KKKKK (K) {20} K

46: KKKKK (K) {20} K

47: KKKKK (K) {20} K

48: KKKKK (K) {20} K

49: KKKKK (K) {20}
50: KKKKK (K) {20}
51: KKKKK (K) {20}
52: KKKKK (K) {20}
53: KKKKK (K) {20}
54: KKKKK (K) {20} X

ABG92627 ck: 9217 len: 68 ! Abg92627 Human DNA-binding protein #53. 11/
(R,K) {20,20}
(K) {20}
38: FLFPE KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
39: LFPEK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
40: FPEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
41: PEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
42: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20} GXXF

ABG92629 ck: 8162 len: 79 ! Abg92629 Human DNA-binding protein #55. 11/
(R,K) {20,20}
(R,K) {20}
41: VRPRV KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
42: RPRVR KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
43: PRVRK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
44: RVRRK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
45: VRKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20} GGRFR

ABG92658 ck: 285 len: 118 ! Abg92658 Human DNA-binding protein #84. 11/
(R,K) {20,20}
(R,K) {20}
98: EKHKQ KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20} G

ABG92659 ck: 5509 len: 58 ! Abg92659 Human DNA-binding protein #85. 11/
(R,K) {20,20}
(K) {20}
36: FYFVC KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
37: YFVCK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
38: FVCKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20} K
39: VCKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

ABG92660 ck: 9074 len: 66 ! Abg92660 Human DNA-binding protein #86. 11/
(R,K) {20,20}
(K) {20}
40: LVQCE KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
41: VQCEK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
42: QCEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
43: CEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
44: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20} KK
46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20} K
47: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

ABG92661 ck: 8528 len: 150 ! Abg92661 Human DNA-binding protein #87. 11/
(R,K) {20,20}
(K) {20}
113: SRNTV KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
114: RNTVK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
115: NTVKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
116: TVKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
117: VKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
118: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
119: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
120: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
121: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20} KXXA

122: KKKKK (K) {20}
123: KKKKK (K) {20} KXXAV
123: KKKKK KKKKKKKKKKKKKKKKKKKKK XHAV

ABG92662 ck: 7676 len: 156 ! Abg92662 Human DNA-binding protein #8. 11/

(R, K) {20, 20}

(K) {20}

108: KLTWI KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

109: TTWIK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

110: TWIKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

111: WIKKK KKKKKKKKKKKKKKKKKKKKK KKKRG

(K) {20}

112: IKKKK KKKKKKKKKKKKKKKKKKKKK KKRGG

(K) {20}

113: KKKKK KKKKKKKKKKKKKKKKKKKKK KRGXG

(K) {20}

114: KKKKK KKKKKKKKKKKKKKKKKKKKK RKGXS

(R, K) {20}

115: KKKKK KKKKKKKKKKKKKKKKKKKKK GKGST

ABG92663 ck: 1736 len: 40 ! Abg92663 Human DNA-binding protein #9. 11/

(R, K) {20, 20}

(K) {20}

18: LPSGL KKKKKKKKKKKKKKKKKKKKK KKK

(K) {20}

19: PGSLL KKKKKKKKKKKKKKKKKKKKK KK

(K) {20}

20: GSLKK KKKKKKKKKKKKKKKKKKKKK K

(K) {20}

21: SLKKK KKKKKKKKKKKKKKKKKKKKK

ABG92665 ck: 1109 len: 98 ! Abg92665 Human DNA-binding protein #91. 11/

(R, K) {20, 20}

(K) {20}

53: QTNNT KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

54: TNNTK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

55: KNTKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

56: NTKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

57: TKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

59: KKKKK (K) {20}
60: KKKKK (K) {20}

60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG

(K) {20}

65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG

(K) {20}

66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGR

(K) {20}

67: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGRS

(K) {20}

68: KKKKK KKKKKKKKKKKKKKKKKKKKK GGRSR

ABG92667 ck: 8102 len: 111 ! Abg92667 Human DNA-binding protein #93. 11/

(R, K) {20, 20}

(K) {20}

78: EFHIL KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

79: FHILK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

80: HILKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

81: ILKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

82: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

87: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

88: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

89: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK

(K) {20}

90: KKKKK KKKKKKKKKKKKKKKKKKKKK KK

1

91: KKKK (K) {20} K
(K) {20}
92: KKKK KKKKKKKKKKKKKKKKK

ABG92668 ck: 8102 len: 111 ! Abg92668 Human DNA-binding protein #94. 11/
(R, K) {20, 20}
(K) {20}

78: EPHIL KKKKKKKKKKKKKKKKK KKKK
(K) {20}

79: PHILK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

80: HILK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

81: ILKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

82: LKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

83: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

84: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

85: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

86: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

87: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

88: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

89: KKKK KKKKKKKKKKKKKKKKK KKK
(K) {20}

90: KKKK KKKKKKKKKKKKKKKKK KK
(K) {20}

91: KKKK KKKKKKKKKKKKKKKKK K
(K) {20}

92: KKKK KKKKKKKKKKKKKKKKK

ABG92669 ck: 8319 len: 53 ! Abg92669 Human DNA-binding protein #95. 11/
(R, K) {20, 20}
(K) {20}

13: RYKRP KKKKKKKKKKKKKKKKK KKKK
(K) {20}

14: YKPK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

15: FKPK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

16: KPKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

17: PKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

1

18: KKKK (K) {20} KKKK
(K) {20}

19: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

20: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

21: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

22: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

23: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

24: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

25: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

26: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

27: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

28: KKKK KKKKKKKKKKKKKKKKK KKEG
(K) {20}

29: KKKK KKKKKKKKKKKKKKKKK KKEG
(K) {20}

30: KKKK KKKKKKKKKKKKKKKKK XEGX
(K) {20}

ABG92671 ck: 7918 len: 80 ! Abg92671 Human DNA-binding protein #97. 11/
(R, K) {20, 20}
(K) {20}

50: NVLTV KKKKKKKKKKKKKKKKK KKKK
(K) {20}

51: VLTVK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

52: LTVK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

53: TVKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

54: VKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

55: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) {20}

56: KKKK KKKKKKKKKKKKKKKKK XXXX
(K) {20}

ABG92673 ck: 4882 len: 41 ! Abg92673 Human DNA-binding protein #99. 11/
(R, K) {20, 20}
(K) {20}

8: FYCFP KKKKKKKKKKKKKKKKK KKKK
(K) {20}

[illegible]

ABG92674 ck: 5469 len: 63 ! Abg92674 Human DNA-binding protein #100. 11

1

[illegible]

```

41: KKKKK KKKKKKKKKKKKKKKKKKKKK KXG
      (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KXG

```

ABG92675 ck: 5075 len: 52 ! Abg92675 Human DNA-binding protein #101. 11

1

```

(R,K){20,20}
(K){20}
30: FIYXK KKKKKKKKKKKKKKKKKKKKK KK
      (K){20}
31: IYVXK KKKKKKKKKKKKKKKKKKKKK KK
      (K){20}
32: VVXXK KKKKKKKKKKKKKKKKKKKKK K
      (K){20}
33: VXXKK KKKKKKKKKKKKKKKKKKKKK
```

ABG92676 ck: 5741 len: 47 ! Abg92676 Human DNA-binding protein #102. 11

1

```

      20: ILTFF XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
              (R,X){20}
      21: LTTFF XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
              (X){20}
      22: TTFFF XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
              (X){20}
      23: TFFFF XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
              (X){20}
      24: FFFFF XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
              (X){20}
      25: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX KXX
              (X){20}
      26: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX KX
              (X){20}
      27: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX X
              (X){20}

```

ABG92677 ck: 2868 len: 84 ! Abg92677 Human DNA-binding protein #103. 11

1

[illegible]

(K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKK

ABG92678 ck: 4686 len: 73 ! Abg92678 Human DNA-binding protein #104. 11

(R,K) {20,20}

(K) {20}

41: YLKE KKKKKKKKKKKKKKKKKKK KKKKK

42: LKKEK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

43: KKEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

44: KEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

45: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K) {20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K) {20}

53: KKKKK KKKKKKKKKKKKKKKKKKK K
(K) {20}

54: KKKKK KKKKKKKKKKKKKKKKKKK

ABG92679 ck: 6676 len: 74 ! Abg92679 Human DNA-binding protein #105. 11

(R,K) {20,20}

(K) {20}

47: LRTFQ KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

48: RTFQK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

49: TFOKK KKKKKKKKKKKKKKKKKKK KKKXX
(K) {20}

50: FOKKK KKKKKKKKKKKKKKKKKKK KKKXX
(K) {20}

51: OKKKK KKKKKKKKKKKKKKKKKKK KXXG
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KXXG

ABG92680 ck: 2283 len: 54 ! Abg92680 Human DNA-binding protein #106. 11

(R,K) {20,20}

(K) {20}

32: IVFCF KKKKKKKKKKKKKKKKKKK KKK

33: VFCFK KKKKKKKKKKKKKKKKKKK KK
(K) {20}

34: PCFKK KKKKKKKKKKKKKKKKKKK X

ABG92681 ck: 7503 len: 74 ! Abg92681 Human DNA-binding protein #107. 11

(R,K) {20,20}

(K) {20}

45: SHLTD KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

46: HLTDK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

47: LTDKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

48: TDKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

49: DKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K) {20}

53: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K) {20}

54: KKKKK KKKKKKKKKKKKKKKKKKK K
(K) {20}

55: KKKKK KKKKKKKKKKKKKKKKKKK

ABG92683 ck: 5199 len: 84 ! Abg92683 Human DNA-binding protein #109. 11

(R,K) {20,20}

(K) {20}

63: AMNAS KKKKKKKKKKKKKKKKKKK XG

ABG92684 ck: 7578 len: 31 ! Abg92684 Human DNA-binding protein #110. 11

(R,K) {20,20}

(K) {20}

6: LTELK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

7: TELKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}


```
8: ELEKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
9: LEKKK KKKKKKKKKKKKKKKKKKKKK KKK
      (K) {20}
10: EKKKK KKKKKKKKKKKKKKKKKKKKK KX
      (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK X
      (K) {20}

ABG92685 ck: 3915 len: 57 1 Abg92685 Human DNA-binding protein #11. 11
      (R,K) {20,20}
      (K) {20}
31: KQLLK KKKKKKKKKKKKKKKKKKKKK KKKKG
      (K) {20}
32: QLLLK KKKKKKKKKKKKKKKKKKKKK KKKGG
      (K) {20}
33: LLLKK KKKKKKKKKKKKKKKKKKKKK KXGGF
      (K) {20}
34: LLKKK KKKKKKKKKKKKKKKKKKKKK XGGF
      (K) {20}

ABG92686 ck: 3679 len: 37 1 Abg92686 Human DNA-binding protein #12. 11
      (R,K) {20,20}
      (K) {20}
15: ISPLT KKKKKKKKKKKKKKKKKKKKK KKK
      (K) {20}
16: SPLTK KKKKKKKKKKKKKKKKKKKKK KX
      (K) {20}
17: PLTKK KKKKKKKKKKKKKKKKKKKKK X
      (K) {20}

ABG92687 ck: 657 len: 196 1 Abg92687 Human DNA-binding protein #13. 11
      (R,K) {20,20}
      (K) {20}
169: FVXFE KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
170: VXFER KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
171: XFERR KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
172: FEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
173: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
174: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
      (K) {20}
175: KKKKK KKKKKKKKKKKKKKKKKKKKK KX
      (K) {20}
176: KKKKK KKKKKKKKKKKKKKKKKKKKK X
      (K) {20}

ABG92688 ck: 4672 len: 57 1 Abg92688 Human DNA-binding protein #14. 11
      (R,K) {20,20}
      (K) {20}
```

```
28: DKTFF KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
29: KTFHK KKKKKKKKKKKKKKKKKKKKK KKKKP
      (K) {20}
30: TFFHK KKKKKKKKKKKKKKKKKKKKK KKKPG
      (K) {20}
31: FHKKK KKKKKKKKKKKKKKKKKKKKK KKKPG
      (K) {20}
32: HKKKK KKKKKKKKKKKKKKKKKKKKK XPGGG
      (K) {20}

ABG92689 ck: 9656 len: 66 1 Abg92689 Human DNA-binding protein #15. 11
      (R,K) {20,20}
      (K) {20}
38: MVISV KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
39: VISVK KKKKKKKKKKKKKKKKKKKKK KKKRE
      (K) {20}
40: ISVKK KKKKKKKKKKKKKKKKKKKKK KKKRE
      (K) {20}
41: SVKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
42: VKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}

ABG92691 ck: 4665 len: 34 1 Abg92691 Human DNA-binding protein #17. 11
      (R,K) {20,20}
      (K) {20}
10: PELLK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
11: ELLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
12: LLLKK KKKKKKKKKKKKKKKKKKKKK KKK
      (K) {20}
13: LLKKK KKKKKKKKKKKKKKKKKKKKK KK
      (K) {20}
14: LKKKK KKKKKKKKKKKKKKKKKKKKK K
      (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK
      (K) {20}

ABG92692 ck: 7810 len: 64 1 Abg92692 Human DNA-binding protein #18. 11
      (R,K) {20,20}
      (K) {20}
37: LKFFW KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
38: KYFWK KKKKKKKKKKKKKKKKKKKKK KKKKG
      (K) {20}
39: YFWKK KKKKKKKKKKKKKKKKKKKKK KKKGX
      (K) {20}
40: FWKKK KKKKKKKKKKKKKKKKKKKKK KKKGP
      (K) {20}
```

41: WKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK XGXP

ABG92693 ck: 269 len: 76 ! Abg92693 Human DNA-binding protein #119. 11

(R, K) {20, 20}

(K) {20}

37: TPSPRA KKKKKKKKKKKKKKKKKKKKK KKKKK

38: PSPRAK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

39: SPRAKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

40: PAKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

41: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKI

(K) {20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKK XKKIK

ABG92694 ck: 8370 len: 45 ! Abg92694 Human DNA-binding protein #120. 11

(R, K) {20, 20}

(K) {20}

17: APKTKQ KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

18: PKTKQK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

19: KTKQKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

20: TKQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

21: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG92695 ck: 1663 len: 87 ! Abg92695 Human DNA-binding protein #121. 11

(R, K) {20, 20}

(K) {20}

36: KWSKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

37: WSSKK KKKKKKKKKKKKKKKKKKKKK KKKKK

38: SSXKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

39: SXKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

40: XKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

41: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

42: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

43: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

44: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

45: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

46: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

47: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

48: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

49: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

50: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

51: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

52: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

53: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

54: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

55: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

56: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

57: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

58: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

59: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

60: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

61: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK KKKKK

62: KKKKK KKKKKKKKKKKKKKKKKKKKKK RGKPF
(R,K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKKK GKPFX

ABG92696 ck: 5607 len: 63 ! Abg92696 Human DNA-binding protein #122. 11

(R,K) {20,20}

(K) {20}

26: MVELE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

27: VELEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

28: ELEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

29: LEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

30: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

31: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

33: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

34: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

35: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKKK GGPF

ABG92697 ck: 5997 len: 58 ! Abg92697 Human DNA-binding protein #123. 11

(R,K) {20,20}

(K) {20}

28: RPTRP KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

29: PTPRP KKKKKKKKKKKKKKKKKKKKKK KKKKK

ABG92698 ck: 5764 len: 74 ! Abg92698 Human DNA-binding protein #124. 11

(R,K) {20,20}

(K) {20}

40: EPLSA KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

41: FLSAK KKKKKKKKKKKKKKKKKKKKKK KKKKK

42: LSARK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

43: SAKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

44: AKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

48: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

50: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

51: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

52: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

53: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

54: KKKKK KKKKKKKKKKKKKKKKKKKKKK X
(K) {20}

ABG92699 ck: 7734 len: 97 ! Abg92699 Human DNA-binding protein #125. 11

(R,K) {20,20}

(K) {20}

57: RGCSY KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

58: GCSYK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

59: CSYKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

60: SYKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

61: YKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

63: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

64: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

65: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

66: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

67: KKKKK (K) {20} KKKKK
 68: KKKKK (K) {20} KKKKK
 69: KKKKK (K) {20} KKKKK
 70: KKKKK (K) {20} KKKKK
 71: KKKKK (K) {20} KKKKK
 72: KKKKK (K) {20} KKKKG
 73: KKKKK (K) {20} KKKGX
 74: KKKKK (K) {20} KKGX
 75: KKKKK (K) {20} XGX

ABG92700 ck: 3164 len: 181 1 Abg92700 Human DNA-binding protein #126. 11

154: TRKPE (R,K) {20,20} KKKKK
 155: RPEPE (K) {20} KKKKK
 156: KPEPE (K) {20} KKKKK
 157: PEPEK (K) {20} KKKKK
 158: EPEKK (K) {20} KKKKK
 159: KPEKK (K) {20} KKKKK
 160: KKKKK (K) {20} KKKKK
 161: KKKKK (K) {20} KKKKK X

ABG92701 ck: 7117 len: 55 1 Abg92701 Human DNA-binding protein #127. 11

22: DDKKN (R,K) {20,20} KKKKK
 23: DKKNK (K) {20} KKKKK
 24: KKNKK (K) {20} KKKKK
 25: KNKKK (K) {20} KKKKK

26: NKKKK (K) {20} KKGGR
 27: KKKKK (K) {20} XGGRF

ABG92702 ck: 9316 len: 67 1 Abg92702 Human DNA-binding protein #128. 11

36: FTLQT (R,K) {20,20} KKKKK
 37: TLQTR (R,K) {20} KKKKK
 38: LQTRK (K) {20} KKKKG
 39: QTRKK (K) {20} KKKGG
 40: TRKKK (K) {20} KKGGR
 41: RKKKK (K) {20} XGGRF

ABG78915 ck: 8085 len: 154 1 Abg78915 Human breast tumour polypeptide #7

114: TQLRQ (R,K) {20,20} KKKKK
 115: QLRQK (K) {20} KKKKK
 116: LRQKK (K) {20} KKKKK
 117: RQKKK (K) {20} KKKKK
 118: QKKKK (K) {20} KKKKK
 119: KKKKK (K) {20} KKKKK
 120: KKKKK (K) {20} KKKKK
 121: KKKKK (K) {20} KKKKK
 122: KKKKK (K) {20} KKKKK
 123: KKKKK (K) {20} KKKKK
 124: KKKKK (K) {20} KKKKK
 125: KKKKK (K) {20} KKKKK
 126: KKKKK (K) {20} KKKKK
 127: KKKKK (K) {20} KKKKK

81: VLKKK (K) {20} KKKKKK
82: LKKKK (K) {20} KKKK
83: KKKKK (K) {20} KKK
84: KKKKK (K) {20} KK
85: KKKKK (K) {20} K
86: KKKKK (K) {20} KKKKKK

ABG65251 ck: 8971 len: 108 ! Abg65251 Human albumin fusion protein #1926

78: TLLXL (R, K) {20, 20} KKKKKK
79: LXLXK (K) {20} KKKKKK
80: LXLKK (K) {20} KKKKKK
81: XLKKK (K) {20} KKKKKK
82: LKKKK (K) {20} KKKKKK
83: KKKKK (K) {20} KKKKKK
84: KKKKK (K) {20} KKKKKK
85: KKKKK (K) {20} KKKKKK
86: KKKKK (K) {20} KKKKKK

ABG65440 ck: 1431 len: 530 ! Abg65440 Human albumin fusion protein #2115

511: LHAP (R, K) {20, 20} KKKKKK
(K) {20}

ABG35896 ck: 5383 len: 86 ! Abg35896 Human peptide encoded by genome-de

15: RRRRG (R, K) {20, 20} KKKKKK
16: RRRGR (R, K) {20} KRRRR
17: RRGRR (R, K) {20} KRRRR
18: RGRRR (R, K) {20} KRRRR
(R, K) {20}

19: GRRRR (R, K) {20} KRRRR
20: RRRRR (R, K) {20} KRRRR
21: RRRRK (R, K) {20} KRRRR
22: RRRKK (R, K) {20} KRRRR
23: RRRKK (R, K) {20} KRRRR
24: RKKKK (R, K) {20} KRRRR
25: KKKKK (R, K) {20} KRRRR
26: KKKKK (R, K) {20} KRRRR
27: KKKKK (R, K) {20} KRRRR
28: KKKKK (R, K) {20} KRRRR
29: KKKKK (R, K) {20} KRRRR
30: KKKKK (R, K) {20} KRRRR
31: RKKKK (R, K) {20} KRRRR
32: KKKKK (R, K) {20} KRRRR
33: KKKKK (R, K) {20} KRRRR
34: KKKKK (R, K) {20} KRRRR
35: KKKKK (R, K) {20} KRRRR
36: KKKKK (R, K) {20} KRRRR
37: RKKKK (R, K) {20} KRRRR
38: KKKKK (R, K) {20} KRRRR
39: KKKKK (R, K) {20} KRRRR
40: KKKRR (R, K) {20} KRRRR
41: KRRRR (R, K) {20} KRRRR
42: RRRRR (R, K) {20} KRRRR
43: RRRRR (R, K) {20} KRRRR

44: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRRR RNRQT
(R) {20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NRQTK

ABG36760 ck: 1334 len: 86 ! Abg36760 Human peptide encoded by genome-de

(R, K) {20, 20}
57: EEEEG (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

58: EEEGR (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

59: EEEGR (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

60: EGRRR (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

61: GRRRK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

62: RRRRK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

63: RRRKK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

64: RRRKK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

65: RRRKK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

66: RRRKK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

67: RRRKK (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

ABG36843 ck: 9082 len: 167 ! Abg36843 Human peptide encoded by genome-de

(R, K) {20, 20}
33: EEEGR (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

34: EGRGR (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

35: GRRGR (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

ABG37280 ck: 2276 len: 89 ! Abg37280 Human peptide encoded by genome-de

(R, K) {20, 20}
23: EEEER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

46: KKEER (K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(K) {20}

47: KKEER (K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(K) {20}

48: EEEER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

1

49: EEEER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

ABG37848 ck: 3607 len: 88 ! Abg37848 Human peptide encoded by genome-de

(R, K) {20, 20}
39: EERRE (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

40: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

41: KREER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

42: RERER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

43: EKKER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

44: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

45: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

46: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

47: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

48: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

49: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

50: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

51: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

52: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

53: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

54: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

55: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

56: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

57: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

58: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

59: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

60: RRRER (R, K) {20} RRRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

```
61: KKKKK (K) {20} KKKKK
62: KKKKK (K) {20} KKKKK
63: KKKKK (K) {20} KKKKK
64: KKKKK (K) {20} KKKKK
65: KKKKK (K) {20} KKKKK
66: KKKKK (K) {20} KKKKK
67: KKKKK (K) {20} KKKKK
68: KKKKK (K) {20} KKKKK
69: KKKKK (K) {20} KKKKK
```

1
ABG38450 ck: 3937 len: 85 1 Abg38450 Human peptide encoded by genome-de

```
1: KKKKK (R,K) {20,20} KKKKK
2: K KKKKK (K) {20} KKKKK
3: KK KKKKK (K) {20} KKKKK
4: KKK KKKKK (K) {20} KKKKK
5: KKKK KKKKK (K) {20} KKKKK
6: KKKKK KKKKK (K) {20} KKKKK
7: KKKKK KKKKK (K) {20} KKKKK
8: KKKKK KKKKK (K) {20} KKKKK
9: KKKKK KKKKK (K) {20} KKKKK
10: KKKKK KKKKK (K) {20} KKKKK
11: KKKKK KKKKK (K) {20} KKKKK
```

ABG39191 ck: 2686 len: 71 1 Abg39191 Human peptide encoded by genome-de

```
20: KKKKK (R,K) {20,20} KKKKK
21: KKKKK (K) {20} KKKKK
```

```
22: KKKKK (K) {20} KKKKK
23: KKKKK (K) {20} KKKKK
24: KKKKK (K) {20} KKKKK
25: KKKKK (R,K) {20} KKKKK
26: KKKKK (R,K) {20} KKKKK
27: KKKKK (R,K) {20} KKKKK
28: KKKKK (R,K) {20} KKKKK
29: KKKKK (R,K) {20} KKKKK
30: KKKKK (R,K) {20} KKKKK
31: KKKKK (R,K) {20} KKKKK
32: KKKKK (R,K) {20} KKKKK
33: KKKKK (R,K) {20} KKKKK
34: KKKKK (R,K) {20} KKKKK
35: KKKKK (R,K) {20} KKKKK
36: KKKKK (R,K) {20} KKKKK
37: KKKKK (R,K) {20} KKKKK
38: KKKKK (R,K) {20} KKKKK
39: KKKKK (R,K) {20} KKKKK
40: KKKKK (R,K) {20} KKKKK
41: KKKKK (R,K) {20} KKKKK
42: KKKKK (R,K) {20} KKKKK
43: KKKKK (R,K) {20} KKKKK
44: KKKKK (R,K) {20} KKKKK
45: KKKKK (R,K) {20} KKKKK
```


46: KKKKK (R,K) {20} KKKSA
47: KKKKK (K) {20} KKSAA
48: KKKKK (K) {20} KSAA
49: KKKKK (K) {20} SAA
ABG40383 ck: 1560 len: 88 ! Abg40383 Human peptide encoded by genome-de
43: RRRRG (R,K) {20,20} RRRRR
44: ERRGR (R) {20} RRRRR
45: RRGRR (R) {20} RRRRR
46: RGRRR (R) {20} RRRRR
47: GRRRR (R) {20} RRRRR
48: RRRRR (R) {20} RRRRR
49: RRRRR (R) {20} RRRRR
50: RRRRR (R) {20} RRRRR
51: RRRRR (R) {20} RRRRR
52: RRRRR (R) {20} RRRRR
53: RRRRR (R) {20} RRRRR
54: RRRRR (R) {20} RRRRR
55: RRRRR (R) {20} RRRRR
56: RRRRR (R) {20} RRRRR
57: RRRRR (R) {20} RRRRR
58: RRRRR (R) {20} RRRRR
59: RRRRR (R) {20} RRRRR
60: RRRRR (R) {20} RRRRR
61: RRRRR (R) {20} RRRRR

62: RRRRR (R) {20} RRRRR
ABG40676 ck: 2324 len: 36 ! Abg40676 Human peptide encoded by genome-de
16: KRRRT (R,K) {20,20} RRRRR
17: ERRTK (R,K) {20} RRRRR
ABG43651 ck: 8343 len: 66 ! Abg43651 Human peptide encoded by genome-de
6: ETERE (R,K) {20,20} KKKKK
7: TEREK (R,K) {20} KKKKK
8: EREKK (R,K) {20} KKKKK
9: REKKK (R,K) {20} KKKKK
ABG45640 ck: 2394 len: 57 ! Abg45640 Human peptide encoded by genome-de
20: EEEEG (R,K) {20,20} RRRRR
21: EEEGR (R) {20} RRRRR
22: EEGRR (R) {20} RRRRR
23: EGRRR (R) {20} RRRRR
24: GRRRR (R) {20} RRRRR
25: RRRRR (R) {20} RRRRR
ABG46635 ck: 3301 len: 52 ! Abg46635 Human peptide encoded by genome-de
12: KKKKK (R,K) {20,20} KKKKK
13: KKKKK (R,K) {20} KKKKK
14: KKKKK (R,K) {20} KKKKK
15: KKKKK (R,K) {20} KKKKK
16: NKKKK (R,K) {20} KKKKK
17: KKKKK (R,K) {20} KKKKK

```

18: KKKKK (R,K){20} KKKKK
19: KKKKK (R,K){20} KKKKK
20: KKKKK (R,K){20} KKKKK
21: KKKKK (R,K){20} KKKKK
22: KKKKK (R,K){20} KKKKK
23: KKKKK (R,K){20} KKKKK
24: KKKKK (R,K){20} KKKKK
25: KKKKK (R,K){20} KKKKK
26: KKKKK (R,K){20} KKKKK
27: KKKKK (R,K){20} KKKKK
28: KKKKK (R,K){20} KKKKK
29: KKKKK (R,K){20} KKKKK
30: KKKKK (R,K){20} KKKKK
31: KKKKK (R,K){20} KKKKK

ABG47062 ck: 4895 len: 51 ! Abg47062 Human peptide encoded by genome-de
(R,K){20,20}
(R,K){20}
18: LKKPM KKKKKKKKKKKKKKKKKKK KKKKK
19: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
20: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
21: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
22: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
23: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
24: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
25: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
26: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
27: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
28: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKPM KKKKKKKKKKKKKKKKKKK KKKKK

ABG7424 ck: 9633 len: 637 ! Abg7424 Human tumour marker protein se2-5.
(R,K){20,20}
(R,K){20}
618: GDKTD KKKKKKKKKKKKKKKKKKK KKKKK

AAE20631 ck: 8971 len: 108 ! Aae20631 Human gene 4 encoded secreted prot
(R,K){20,20}
(R,K){20}
78: TLXL KKKKKKKKKKKKKKKKKKK KKKKK
79: LXLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
(K){20}

```

```

80: LVLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
81: XLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAE20642 ck: 7046 len: 105 ! Aae20642 Human gene 4 encoded secreted prote
      (R, K) {20, 20}
      (K) {20}
78: TLVLK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
79: LVLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
80: LVLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
81: VLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKK K
      (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKK

ABB9690 ck: 1158 len: 226 ! Abb9690 Human polypeptide SHQ ID NO 2066.
      (R, K) {20, 20}
      (K) {20}
198: GAESL KKKKKKKKKKKKKKKKKKKKK KGGPP
      (K) {20}
199: AESLK KKKKKKKKKKKKKKKKKKKKK KGGPP
      (K) {20}
200: ESLKK KKKKKKKKKKKKKKKKKKKKK GPPXX

AAE14544 ck: 7250 len: 35 ! Aae14544 Peptide p65 used in assay for detec
      (R, K) {20, 20}
      (K) {20}
1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK

```

```

3:  KK (K) {20}
   KK (K) {20}
4:  KK (K) {20}
   KK (K) {20}
5:  KK (K) {20}
   KK (K) {20}
6:  KK (K) {20}
   KK (K) {20}
7:  KK (K) {20}
   KK (K) {20}
8:  KK (K) {20}
   KK (K) {20}
9:  KK (K) {20}
   KK (K) {20}
10: KK (K) {20}
   KK (K) {20}
11: KK (K) {20}
   KK (K) {20}
12: KK (K) {20}
   KK (K) {20}
13: KK (K) {20}
   KK (K) {20}
14: KK (K) {20}
   KK (K) {20}
15: KK (K) {20}
   KK (K) {20}
16: KK (K) {20}
   KK (K) {20}

```

AAU75162 ck: 5536 len: 84 1 AAU75162 Single-chain antigen-binding polyP

```

1:  (R,K) {20,20}
   (K) {20}
2:  K (K) {20}
   KK (K) {20}
3:  KK (K) {20}
   KK (K) {20}
4:  KK (K) {20}
   KK (K) {20}
5:  KK (K) {20}
   KK (K) {20}
6:  KK (K) {20}
   KK (K) {20}
7:  KK (K) {20}
   KK (K) {20}
8:  KK (K) {20}
   KK (K) {20}
9:  KK (K) {20}
   KK (K) {20}

```

```

10: KK (K) {20}
   KK (K) {20}
11: KK (K) {20}
   KK (K) {20}
12: KK (K) {20}
   KK (K) {20}
13: KK (K) {20}
   KK (K) {20}
14: KK (K) {20}
   KK (K) {20}
15: KK (K) {20}
   KK (K) {20}
16: KK (K) {20}
   KK (K) {20}
17: KK (K) {20}
   KK (K) {20}
18: KK (K) {20}
   KK (K) {20}
19: KK (K) {20}
   KK (K) {20}
20: KK (K) {20}
   KK (K) {20}
21: KK (K) {20}
   KK (K) {20}
22: KK (K) {20}
   KK (K) {20}
23: KK (K) {20}
   KK (K) {20}
24: KK (K) {20}
   KK (K) {20}
25: KK (K) {20}
   KK (K) {20}
26: KK (K) {20}
   KK (K) {20}
27: KK (K) {20}
   KK (K) {20}
28: KK (K) {20}
   KK (K) {20}
29: KK (K) {20}
   KK (K) {20}
30: KK (K) {20}
   KK (K) {20}
31: KK (K) {20}
   KK (K) {20}
32: KK (K) {20}
   KK (K) {20}
33: KK (K) {20}
   KK (K) {20}

```

34: KKKKK (K) {20} KKKKK
 35: KKKKK (K) {20} KKKKK
 36: KKKKK (K) {20} KKKKK
 37: KKKKK (K) {20} KKKKK

AAU75163 ck: 8024 len: 84 ! Aau75163 Single-chain antigen-binding poly

1 (R,K) {20,20} RRRRR
 1: RRRRR
 2: R (R) {20} RRRRR
 3: R (R) {20} RRRRR
 4: R (R) {20} RRRRR
 5: R (R) {20} RRRRR
 6: R (R) {20} RRRRR
 7: R (R) {20} RRRRR
 8: R (R) {20} RRRRR
 9: R (R) {20} RRRRR
 10: R (R) {20} RRRRR
 11: R (R) {20} RRRRR
 12: R (R) {20} RRRRR
 13: R (R) {20} RRRRR
 14: R (R) {20} RRRRR
 15: R (R) {20} RRRRR
 16: R (R) {20} RRRRR
 17: R (R) {20} RRRRR
 18: R (R) {20} RRRRR
 19: R (R) {20} RRRRR

20: R (R) {20} RRRRR
 21: R (R) {20} RRRRR
 22: R (R) {20} RRRRR
 23: R (R) {20} RRRRR
 24: R (R) {20} RRRRR
 25: R (R) {20} RRRRR
 26: R (R) {20} RRRRR
 27: R (R) {20} RRRRR
 28: R (R) {20} RRRRR
 29: R (R) {20} RRRRR
 30: R (R) {20} RRRRR
 31: R (R) {20} RRRRR
 32: R (R) {20} RRRRR
 33: R (R) {20} RRRRR
 34: R (R) {20} RRRRR
 35: R (R) {20} RRRRR
 36: R (R) {20} RRRRR
 37: R (R) {20} RRRRR

AAU75164 ck: 9643 len: 83 ! Aau75164 Single-chain antigen-binding poly

1 (R,K) {20,20} RRRRR
 1: RRRRR
 2: R (R,K) {20} RRRRR
 3: R (R,K) {20} RRRRR
 4: R (R,K) {20} RRRRR
 5: R (R,K) {20} RRRRR

6: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
7: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
8: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
9: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
10: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
11: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
12: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
13: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
14: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
15: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
16: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
17: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
18: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
19: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
20: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
21: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
22: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
23: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
24: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
25: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
26: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
27: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
28: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
29: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}

1

30: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
31: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
32: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
33: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
34: RKKR KRKKRKRKKRKRKKR KRKK
(R,K) {20}
35: KRKR KRKKRKRKKRKRKKR RKKR
(R,K) {20}
36: RKKR KRKKRKRKKRKRKKR KKKK
(R,K) {20}
37: KRKR KRKKRKRKKRKRKKR XXXX
(R,K) {20}
AAU75165 ck: 6708 len: 84 ! AAU75165 single-chain antigen-binding polyp
(R,K) {20,20}
1: RRRRRRRRRRRRRRRRR RRR
(R) {20}
2: R RRRRRRRRRRRRRRR RRR
(R) {20}
3: RR RRRRRRRRRRRRRRR RRR
(R) {20}
4: RRR RRRRRRRRRRRRRRR RRR
(R) {20}
5: RRR RRRRRRRRRRRRRRR RRR
(R) {20}
6: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
7: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
8: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
9: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
10: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
11: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
12: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
13: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
14: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}
15: RRRR RRRRRRRRRRRRRRR RRR
(R) {20}

16: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
17: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
18: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
19: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
20: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
21: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
22: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
23: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
24: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
26: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
27: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
28: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
29: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
30: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
31: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
32: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
33: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
34: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
35: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
36: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
37: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}

2: K RKRKRRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 3: KR KRKRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 4: KRK RKRRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 5: KRKR KRKRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 6: KRKRK RKRRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 7: RKRKR KRKRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 8: KRKRK RKRRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 9: RKRKR KRKRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 10: KRKRK RKRRKRKRKRKRKRKRKR KRKR
 (R,K) {20}
 11: RKRKR KRKRKRKRKRKRKRKRKR KR
 (R,K) {20}
 12: KRKRK RKRRKRKRKRKRKRKRKRKR R
 (R,K) {20}
 13: RKRKR KRKRKRKRKRKRKRKRKRKR
 (R,K) {20}

AAm48798 ck: 6221 len: 26 | Aam48798 Tumour-targeting peptide vector pE

1: (R,K) {20,20}
 (K) {20}

7: RGDCF KKKKKKKKKKKKKKKKKKK

ABB44830 ck: 5275 len: 38 | Abb44830 Human protective sequence CNI-00745

1: (R,K) {20,20}
 (K) {20}

4: MAQ KKKKKKKKKKKKKKKKKKK KKKKK

5: MAQK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

6: MAQK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

7: AQKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

8: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

13: KKKKK (K) {20}
14: KKKKK (K) {20}
15: KKKKK (K) {20}
16: KKKKK (K) {20}
17: KKKKK (K) {20} GG

1

AAU69736 ck: 875 len: 46 1 Aau69736 Cell death protective sequence CNT
(R, K) {20, 20}
15: RKKKS KKKKK (K) {20}
16: EKKSK KKKKK (K) {20}
17: KKSJK KKKKK (K) {20}
18: KSKKK KKKKK (K) {20}
19: SKKKK KKKKK (K) {20}
20: KKKKK KKKKK (K) {20}
21: KKKKK KKKKK (K) {20}
22: KKKKK KKKKK (K) {20}
23: KKKKK KKKKK (K) {20}
24: KKKKK KKKKK (K) {20} KGG
25: KKKKK KKKKK (K) {20} GG

1

AAU69736 ck: 5862 len: 50 1 Aau69736 Cell death protective sequence CNT
(R, K) {20, 20}
11: ESALG RKKKK (R, K) {20}
12: SALGR KKKKK (K) {20}
13: ALGRK KKKKK (K) {20}
14: LGRKK KKKKK (K) {20}
15: GRKKK KKKKK (K) {20}
16: RKKKK KKKKK (K) {20}

1

17: KKKKK (K) {20}
18: KKKKK (K) {20}
19: KKKKK (K) {20}
20: KKKKK (K) {20}
21: KKKKK (K) {20}
22: KKKKK (K) {20}
23: KKKKK (K) {20}
24: KKKKK (K) {20}
25: KKKKK (K) {20}
26: KKKKK (K) {20}
27: KKKKK (K) {20}
28: KKKKK (K) {20} KGR
29: KKKKK (K) {20} GR

ABG47914 ck: 5383 len: 86 1 Abg47914 Human liver peptide, SEQ ID No 265
(R, K) {20, 20}
15: RRRRG RRRKK (R, K) {20} KRRR
16: RRRGR RRRKK (R, K) {20} KRRR
17: RRRGR RRRKK (R, K) {20} RRRR
18: RRRRR RRRKK (R, K) {20} RRRR
19: GRRRR RRRKK (R, K) {20} RRRR
20: RRRRR RRRKK (R, K) {20} RRRR
21: RRRRK RRRKK (R, K) {20} RRRR
22: RRRKK RRRKK (R, K) {20} RRRR
23: RRRKK RRRKK (R, K) {20} RRRR
24: RKKKK RRRKK (R, K) {20} RRRR

[illegible][illegible]

```

ABG48847 ck: 9082 len: 167 ! Abg48847 Human liver peptide, SEQ ID No 274:
      (R, K){20,20}
      (R, K){20}
33: EEGRG RRRRRRRRRRRRRRRRRR RRGCG
      (R, K){20}
34: EGRGR RRRRRRRRRRRRRRRRRR RGGGR
      (R, K){20}
35: GRGRR RRRRRRRRRRRRRRRRRR GGGRR

```

```

ABG49354 ck: 2276 len: 89 | Abg49354 Human liver peptide, SEQ ID No 280
      (R, K) {20, 20}
23: EEEEE KKKKKRKRKKKKRKKKK EEEKK
      (R, K) {20}
45: KKEE KKKKKKKKKKKKKKKKK KKEE
      (K) {20}
47: KEEK KKKKKKKKKKKKKKKKK RKEE
      (K) {20}
48: EEEK KKKKKKKKKKKKKKKKK KEEE
      (R, K) {20}
49: EEEK KKKKKKKKKKKKKKKKK EEEEE

```

```

ABG49651 ck: 1939 len: 130 ! Abg49651 Human liver peptide, SEQ ID No 282:
(R,K){20,20}
(R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRK RRRR
(R,K){20}
43: GRKER RRRRRRRRRRRRRRRRRK RRRR
(R,K){20}
44: RKEER RRRRRRRRRRRRRRRRRK RRRR
(R,K){20}

```


1

```
45: KERRR RRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
46: ERRRR RRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
47: RRRRR RRRRRRRRRRRRRRRRRR RKKKK (R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRR RKKKK (R,K){20}
49: RRRRR RRRRRRRRRRRRRRRRRR RKKKK (R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRR RKKKE (R,K){20}
51: RRRRR RRRRRRRRRRRRRRRRRR RKKKE (R,K){20}
52: RRRRR RRRRRRRRRRRRRRRRRR RKKEE (R,K){20}
53: RRRRR RRRRRRRRRRRRRRRRRR RKEEE (R,K){20}
54: RRRRR RRRRRRRRRRRRRRRRRR EEEEE (R,K){20}

ABG4962 ck: 3607 len: 88 ! Abg4962 Human liver peptide, SEQ ID No 286
(R,K){20,20}
39: ERRRE RRRRRRRRRRRRRRRRRR RKKKK (K){20}
40: RKREK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
41: KREKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
42: REKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
43: EKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
44: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
45: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
46: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
47: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
48: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
49: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
50: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
51: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
```

1

```
52: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
53: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
54: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
55: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
56: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
57: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
58: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
59: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
60: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
61: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
62: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
63: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
64: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
65: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
66: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
67: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
68: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}
69: KKKKK RRRRRRRRRRRRRRRRRR RKKKK (K){20}

ABG50530 ck: 3937 len: 85 ! Abg50530 Human liver peptide, SEQ ID No 291
(R,K){20,20}
1: KRRRRRRRRRRRRRRRRR RKKKK (K){20}
2: KRRRRRRRRRRRRRRRRR RKKKK (K){20}
3: KRRRRRRRRRRRRRRRRR RKKKK (K){20}
4: KRRRRRRRRRRRRRRRRR RKKKK (K){20}
5: KRRRRRRRRRRRRRRRRR RKKKK (K){20}
```


1
ABG59254 ck: 3301 len: 52 ! Abg59254 Human liver peptide, SEQ ID No 379
(R,K){20,20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
ABG59685 ck: 4895 len: 51 ! Abg59685 Human liver peptide, SEQ ID No 383
(R,K){20,20}
18: LFKPM KKKKKKKKKKKKKKKKKRR KKLTT
(R,K){20}
19: FKPMR KKKKKKKKKKKKKKKKKRR KLT TT

1
ABB96057 ck: 5626 len: 139 ! Abb96057 Human testicular antigen SEQ ID NC
(R,K){20,20}
111: IHLNL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
112: HLNLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
113: LNLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
114: NLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
1
ABB96575 ck: 4751 len: 80 ! Abb96575 Human testicular antigen SEQ ID NC
(R,K){20,20}
(K){20}
61: KKKFD KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
1
AAU87124 ck: 5599 len: 281 ! Aau87124 Novel central nervous system prote
(R,K){20,20}
(K){20}
248: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
249: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
250: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
251: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
252: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
253: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
254: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
255: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
256: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
257: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
258: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
259: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
260: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
261: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

262: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKK

ABG00401 ck: 5116 len: 1,074 | Abg00401 Novel human diagnostic protein #39

(R, K) {20, 20}

(R, K) {20}

609: RGSSS KKKKKKKKKKKKKKKKKRR KNRKK

(R, K) {20}

610: GSSSK KKKKKKKKKKKKKKKRRK NRKKK

ABG03974 ck: 5341 len: 99 | Abg03974 Novel human diagnostic protein #39

(R, K) {20, 20}

(K) {20}

2: M KKKKKKKKKKKKKKKKKKK KKKKN

(K) {20}

3: MK KKKKKKKKKKKKKKKKKKK KKKNS

(K) {20}

4: MKK KKKKKKKKKKKKKKKKKKK KKSQ

(K) {20}

5: MKKK KKKKKKKKKKKKKKKKKKK KNSOI

(K) {20}

6: MKKKK KKKKKKKKKKKKKKKKKKK NSQID

ABG04391 ck: 9047 len: 139 | Abg04391 Novel human diagnostic protein #43

(R, K) {20, 20}

(K) {20}

79: EEEEE KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

80: EEEEE KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

81: EEEEE KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

82: EEEEE KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

83: EEEEE KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

89: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

90: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

91: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

92: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

93: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

94: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

95: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

96: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

97: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

98: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

99: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

107: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

108: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

109: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

110: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

111: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

112: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

113: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

ABG05352 ck: 1276 len: 204 | Abg05352 Novel human diagnostic protein #53'

(R, K) {20, 20}

(K) {20}

```

106: EEEEE KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
107: EEEEE KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
108: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
109: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
110: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
111: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
112: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
113: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
114: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
115: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
116: KKKKK KKKKKKKKKKKKKKKKKKKKK KKEE
      (R,K) {20}
117: KKKKK KKKKKKKKKKKKKKKKKKKKK KKEE
      (R,K) {20}
118: KKKKK KKKKKKKKKKKKKKKKKKKKK EEEK
      (R,K) {20}

ABG05367 ck: 6907 len: 808 1 Abg05367 Novel human diagnostic protein #53
      (R,K) {20,20}
      (R,K) {20}
219: EEEEE RKKKKKKKKKKKKKKKKKKRR RRRR
      (R,K) {20}
220: EEEEE RKKKKKKKKKKKKKKKKRRR RRRR
      (R,K) {20}
221: EEEEE KKKKKKKKKKKKKKKKKRRR RRRR
      (R,K) {20}
222: EERRR KKKKKKKKKKKKKKKKKRRR RRRK
      (R,K) {20}
223: ERRKK KKKKKKKKKKKKKKKKKRRR RRRK
      (R,K) {20}
224: RRRKK KKKKKKKKKKKKKKKKKRRR RRRK
      (R,K) {20}
225: RKKKK KKKKKKKKKKKKKKKKKRRR KKKK
      (R,K) {20}
226: KKKKK KKKKKKKKKKKKKKKKKRRR KKKK
      (R,K) {20}
227: KKKKK KKKKKKKKKKKKKKKKKRRK KKKK
      (R,K) {20}
228: KKKKK KKKKKKKKKKKKKKKKKRRK KKKK
      (R,K) {20}

```

```

229: KKKKK RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
230: KKKKK RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
231: KKKRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
232: KKKRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
233: KRRRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
234: RRRRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
235: RRRRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
236: RRRRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
237: RRRRR RRRRRRRRRRRRRRRRRKKK KKKK
      (R,K) {20}
258: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
259: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}
260: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
      (K) {20}

ABG06375 ck: 7807 len: 2,570 1 Abg06375 Novel human diagnostic protein #63
      (R,K) {20,20}
      (R,K) {20}
205: KKKSS RRRRRRRRRRRRRRRRRKKK RRRQ
      (R,K) {20}
206: KSSRR RRRRRRRRRRRRRRRRRKKK RRRQ
      (R,K) {20}
207: KSSRR RRRRRRRRRRRRRRRRRKKK RRRQ
      (R,K) {20}
208: SSRRR RRRRRRRRRRRRRRRRRKKK RRRQ
      (R,K) {20}

ABG06513 ck: 2934 len: 154 1 Abg06513 Novel human diagnostic protein #65
      (R,K) {20,20}
      (R,K) {20}
49: KKKKT RKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
50: KKKTR KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
51: KKTTR KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
52: KTRKK RKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
53: TRKKR KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}
54: RKKRK KKKKKKKKKKKKKKKKKKKKK KKKK
      (R,K) {20}

```

(R,K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKK
SEXDR
ABG07742 ck: 8672 len: 502 1 Abg07742 Novel human diagnostic protein #77
(R,K){20,20}
54: DDEEE RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
55: DEEER RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
56: EEEER RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
57: EEEEE RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
58: EEEEE RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
59: RRRKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
60: RRRKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
61: RKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKK
KNIL

(R,K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKK
NILYR
ABG10052 ck: 7107 len: 39 1 Abg10052 Novel human diagnostic protein #10
(R,K){20,20}
2: Q KKKKKKKKKKKKKKKKKKK
(K){20}
3: QK KKKKKKKKKKKKKKKKKKK
(K){20}
4: QKK KKKKKKKKKKKKKKKKKKK
(K){20}
5: QKK KKKKKKKKKKKKKKKKKKK
(K){20}
6: QKKK KKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK
EKKKK
ABG10053 ck: 3274 len: 189 1 Abg10053 Novel human diagnostic protein #10
(R,K){20,20}
70: EKEKE KKKKKKKKKKKKKKKKKKK
(R,K){20}
71: KEKEK KKKKKKKKKKKKKKKKKKK
(R,K){20}
72: EKEKK RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
73: KEKKR KKKKKKKKKKKKKKKKKKK
(R,K){20}
74: EKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
76: KKKKK RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
77: RKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKK
EEEE
ABG11241 ck: 3870 len: 121 1 Abg11241 Novel human diagnostic protein #11
(R,K){20,20}
44: KKKKE KKKKKKKKKKKKKKKKKKK
(K){20}
45: KKEEK KKKKKKKKKKKKKKKKKKK
(K){20}

46: KEEKK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K) {20}
47: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K) {20}
48: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKE

ABG11242 ck: 4493 len: 100 ! Abg11242 Novel human diagnostic protein #11

1

(R,K) {20,20}
42: ETPSE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
43: TPSEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
44: PSEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
45: SEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
46: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

1

61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG11245 ck: 2517 len: 85 ! Abg11245 Novel human diagnostic protein #11

(R,K) {20,20}
33: EEEEE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG11250 ck: 4343 len: 92 1 Abg11250 Novel human diagnostic protein #11

 (R,K) {20,20}
 (K) {20}
53: EKEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
54: KEKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
55: EKEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
56: KEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
57: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG11266 ck: 9563 len: 146 1 Abg11266 Novel human diagnostic protein #11

 (R,K) {20,20}
 (K) {20}
36: KEKKE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
37: EKREK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
38: KREKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
39: REKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
40: EKKKK KKKKKKKKKKKKKKKKKKKKK EEEEE
 (R,K) {20}
79: EEEEE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
80: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK

 (R,K) {20}
81: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
82: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
83: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
91: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
93: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG11277 ck: 8026 len: 1,080 1 Abg11277 Novel human diagnostic protein #11:

 (R,K) {20,20}
 (R,K) {20}
709: QKEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
710: KEKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
711: EKEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
712: KEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
713: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
714: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
715: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
716: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
717: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R,K) {20}
718: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

719: KKKK (R,K){20} KKKK
 720: KKKK (R,K){20} KKKK
 721: KKKK (R,K){20} KKKK
 722: KKKK (R,K){20} KKKK
 723: KKKK (R,K){20} KKKK
 724: KKKK (R,K){20} KKKK
 725: KKKK (R,K){20} KKKK
 726: KKKK (R,K){20} KKKK
 727: KKKK (R,K){20} EERRK

ABG11732 ck: 2886 len: 56 1 Abg11732 Novel human diagnostic protein #11
 8: KRRG (R,K){20,20} RRRR
 9: RRRG (R){20} RRRR
 10: RRRG (R){20} RRRR
 11: RRRR (R){20} RRRR
 12: GRRR (R){20} RRRR
 13: RRRR (R){20} RRRR

1
 ABG11734 ck: 4548 len: 142 1 Abg11734 Novel human diagnostic protein #11
 109: QMLSV (R,K){20,20} KKKK
 110: MLSV (R,K){20} KKKK
 111: LSVK (R,K){20} KKKK
 112: SVKK (R,K){20} KKKK
 113: VKKK (R,K){20} KKKK
 114: KKKK (R,K){20} KKKK
 115: KKKK (K){20} KKKK

116: KKKK (K){20} KKKK
 117: KKKK (K){20} KKKK
 118: KKKK (K){20} KKKK

ABG11738 ck: 3009 len: 567 1 Abg11738 Novel human diagnostic protein #11
 526: EEEE (R,K){20,20} KKKK
 527: EEEER (R,K){20} KKKK
 528: EEEER (R,K){20} KKKK
 529: EERRK (K){20} KKKK
 530: EERRK (K){20} KKKK
 531: RRRK (K){20} KKKK
 532: RRRK (K){20} KKKK
 533: KKKK (R,K){20} KKKK
 534: KKKK (R,K){20} EERRA

1
 ABG22512 ck: 8641 len: 856 1 Abg22512 Novel human diagnostic protein #22
 130: KKKF (R,K){20,20} KKKK
 131: KKKF (K){20} KKKK
 132: KKKF (K){20} KKKK
 133: KKKK (K){20} KKKK
 134: PKKK (K){20} KKKK
 135: KKKK (K){20} KKKK
 136: KKKK (K){20} QSLV
 ABG22638 ck: 4197 len: 896 1 Abg22638 Novel human diagnostic protein #22
 227: GREER (R,K){20,20} RRRR

1

82: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KGNLSS
83: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK GNLSSV

ABG26492 ck: 5234 len: 68 ! Abg26492 Novel human diagnostic protein #26

(R,K) {20,20}

29: KKKEE (R) {20} RRRRRRRRRRRRRRRRRRR RRRRG
(R) {20}

30: KKEER (R) {20} RRRRRRRRRRRRRRRRRRR RRRGK
(R) {20}

31: KEERR (R) {20} RRRRRRRRRRRRRRRRRRR RRGKK
(R) {20}

32: EERRR (R) {20} RRRRRRRRRRRRRRRRRRR RGKGD
(R) {20}

33: ERRRR (R) {20} RRRRRRRRRRRRRRRRRRR GKKG

ABG26493 ck: 4204 len: 80 ! Abg26493 Novel human diagnostic protein #26

(R,K) {20,20}

27: EKEKE (R,K) {20} KRRRRRRRRRRRRRRRRRR RRRR
(R,K) {20}

28: KEKEK (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

29: EKEK (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

30: KEKRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

31: EKRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

32: KRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

33: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

34: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

35: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

36: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

37: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

38: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

39: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

40: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

41: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

1

42: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

43: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

44: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

45: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

46: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

47: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

48: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

49: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

50: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRRT
(R) {20}

51: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRNT
(R) {20}

52: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRNN
(R) {20}

53: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRNE
(R) {20}

ABG26496 ck: 8829 len: 90 ! Abg26496 Novel human diagnostic protein #26

(R,K) {20,20}

61: EKKEE (R,K) {20} KRRRRRRRRRRRRRRRRRR RRRR
(R,K) {20}

62: KKEEK (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

63: KEKRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

64: EEKRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

65: EKRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

66: KRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRR
(R) {20}

67: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRN
(R) {20}

68: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRN
(R) {20}

69: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRN
(R) {20}

70: RRRRR (R) {20} RRRRRRRRRRRRRRRRRRR RRRN
(R) {20}

ABG26497 ck: 9186 len: 115 ! Abg26497 Novel human diagnostic protein #26

72: RRRR {R} {20} RRRR
73: RRRR {R} {20} RRRR
74: RRRR {R} {20} RRRR
75: RRRR {R} {20} RRRR
76: RRRR {R} {20} RRRR
77: RRRR {R} {20} RRRR
78: RRRR {R} {20} RRRR
79: RRRR {R} {20} RRRR
80: RRRR {R} {20} RRRR
81: RRRR {R} {20} RRRR
82: RRRR {R} {20} RRRR
83: RRRR {R} {20} RRRR
84: RRRR {R} {20} RRRR
85: RRRR {R} {20} RRRR
86: RRRR {R} {20} RRRR

ABG26498 ck: 7156 len: 140 1 Abg26498 Novel human diagnostic protein #20

(R, K) {20, 20}
97: RRGGG RRRRRRRRRRRRRRRRRRRRR
(R) {20}
98: RCGGR RRARRRRRRRRRRRRRRRRRRRR
(R) {20}
99: GGGRR RRRRRRRRRRRRRRRRRRRRR
(R) {20}
100: GGRRR RRARRRRRRRRRRRRRRRRRRRR
(R) {20}
101: GRRRR RRRRRRRRRRRRRRRRRRRRR
(R) {20}
102: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R) {20}
103: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R) {20}
104: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R) {20}

```

105: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
106: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
107: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
108: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
109: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
110: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
111: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
112: RRRRR (R) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
113: RRRRR (R, K) {20} RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

```

[illegible]

1

AB026501 ck: 1730 len: 182 i Ab026501 Novel human diagnostic protein #26

(R,K){20,20}

(K){20}

72: KEEKE KKKKKKKKKKKKKKKKKKK KKKKE

(K){20}

73: EKEKE KKKKKKKKKKKKKKKKKKK KKKKE

(K){20}

74: EKEKE KKKKKKKKKKKKKKKKKKK KKEEE

(K){20}

75: KEKKE KKKKKKKKKKKKKKKKKKK KEEEE

(K){20}

76: EKKEK KKKKKKKKKKKKKKKKKKK EEEEA

(R){20}

133: GRRRS RRRRRRRRRRRRRRRRRR RRRR

(R){20}

134: RRRSR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

135: RRSRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

136: RSRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

137: SRRSR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

138: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

139: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

140: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

141: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

142: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

143: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

144: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

145: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

146: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

147: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

148: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R){20}

149: RRRRR RRRRRRRRRRRRRRRRRR RRRR

(R,K){20}

150: RRRRR RRRRRRRRRRRRRRRRRR RRRR

1
ABG26502 ck: 3399 len: 101 ! Abg26502 Novel human diagnostic protein #26
(R,K){20,20}
59: EEEEG RRRRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
60: EEEGR RRRRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
61: EEEGR RRRRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
62: EGRRR RRRKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
63: GRRRR RRRKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
64: RRRRR RKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
65: RRRRR RKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
66: RRRRK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
67: RRRKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
68: RKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
69: RKKKK KKKKKKKKKKKKKKKKKKK EYFQT
(R,K){20}

1
ABG26505 ck: 4704 len: 93 ! Abg26505 Novel human diagnostic protein #26
(R,K){20,20}
60: KEEEG RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
61: EEEGR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
62: EEEGR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
63: EGRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
64: GRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R){20}
65: RRRRR RRRRRRRRRRRRRRRRRR RKKKK
(R){20}
66: RRRRR RRRRRRRRRRRRRRRRRR KKKKK
(R){20}
67: RRRRR RRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
68: RRRRR RRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
69: RRRRR RRRRRRRRRRRRRRRRRR EEEEE
(R,K){20}

ABG26506 ck: 3684 len: 85 ! Abg26506 Novel human diagnostic protein #26

1

```
(R,K){20,20}
47: EKEKE KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
48: KEKEK RKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
49: EKEKR KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
50: KEKKR RKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
51: EKRRR KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
52: KRKRK RKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
53: RKRKR KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
54: KRKRK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
55: RKRKR KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
56: KRKRK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
57: RKRKR KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
ABG26507 ck: 9838 len: 109 1 Abg26507 Novel human diagnostic protein #26
(R,K){20,20}
60: EEEEE RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
61: EEEEE RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
62: EEEEE RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
63: EEEEE RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
64: EEEEE RRRRRRRRRRRRRRRRRRRRRRRRR
(R,K){20}
65: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRR
(R,K){20}
```

1

```
ABG26508 ck: 4488 len: 121 1 Abg26508 Novel human diagnostic protein #26
(R,K){20,20}
56: DDEEB RRRKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
57: DEEBR RRRKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
58: EEERR RKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
59: EERRR KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
60: EERRR KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
61: RRRKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
62: RRRKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
63: RKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
ABG26510 ck: 3426 len: 74 1 Abg26510 Novel human diagnostic protein #26
(R,K){20,20}
47: RRRRS RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
48: RRRSR RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
49: RRSRR RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
50: RSRRR RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
51: SRRRR RRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
ABG26513 ck: 3117 len: 265 1 Abg26513 Novel human diagnostic protein #26
(R,K){20,20}
199: EEEEE KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
222: KEEEE KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
```

223: KEEBK KKKKKKKKKKKKKKKKK RKEEB
(R,K){20}
224: EEEKK KKKKKKKKKKKKKKKKK KEEEB
(R,K){20}
225: EEEKK KKKKKKKKKKKKKKKKK EEEEB

ABG26514 ck: 9106 len: 218 | Abg26514 Novel human diagnostic protein #26

(R,K){20,20}
182: EEEBK RRRRRRRRRRRRRRRRR RRRR
(R){20}
183: EERGR RRRRRRRRRRRRRRRRR RRRR
(R){20}
184: RRRGR RRRRRRRRRRRRRRRRR RRRR
(R){20}
185: RGRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
186: GRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
187: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
188: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
189: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
190: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
191: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
192: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
193: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}

ABG26515 ck: 7288 len: 389 | Abg26515 Novel human diagnostic protein #26

(R,K){20,20}
259: KKEES RKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
260: KEESR KKKKKKKKKKKKKKKKK KKKKK
(K){20}
261: EESRK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
262: ESRKK KKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
263: SRKKK KKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
264: RKKKK KKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

265: KKKKK KKKKKKKKKKKKKKKKK NKKKK

ABG26516 ck: 2295 len: 91 | Abg26516 Novel human diagnostic protein #26

(R,K){20,20}
45: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
46: EEEER RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
47: EEEER RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
48: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
49: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
51: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
52: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
53: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
54: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}

ABG26518 ck: 8431 len: 761 | Abg26518 Novel human diagnostic protein #26

(R,K){20,20}
725: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R){20}
726: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R){20}
727: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R){20}
728: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R){20}
729: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R){20}
730: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
731: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
732: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
733: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
734: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
735: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}

76: RRRRR RRRRRRRRRRRRRRRRRR RKRI
(R) {20}
77: RRRRR RRRRRRRRRRRRRRRRRR RKI
(R) {20}
78: RRRRR RRRRRRRRRRRRRRRRRR KI
(R,K) {20}
79: RRRRR RRRRRRRRRRRRRRRRRR I

ABg26521 ck: 482 len: 367 ! Abg26521 Novel human diagnostic protein #26

(R,K) {20,20}

(K) {20}

317: EEEEG KKKKKKKKKKKKKKKKKKK KKKK

318: EEEGK KKKKKKKKKKKKKKKKKKK KKKR
(K) {20}

319: EEEGK KKKKKKKKKKKKKKKKKKK KKKR
(K) {20}

320: EGGKK KKKKKKKKKKKKKKKKKKK KRRK
(K) {20}

321: GKKKK KKKKKKKKKKKKKKKKKKK RRRK
(R,K) {20}

322: KKKKK KKKKKKKKKKKKKKKKKKK RRRK
(R,K) {20}

323: KKKKK KKKKKKKKKKKKKKKKKKK RKKK
(R,K) {20}

324: KKKKK KKKKKKKKKKKKKKKKKKK RKKK
(R,K) {20}

325: KKKKK KKKKKKKKKKKKKKKKKKK RKKK
(R,K) {20}

326: KKKKK KKKKKKKKKKKKKKKKKKK KKEQ
(R,K) {20}

327: KKKKK KKKKKKKKKKKKKKKKKKK KEGQ
(R,K) {20}

328: KKKKK KKKKKKKKKKKKKKKKKKK EQQL
(R,K) {20}

ABg26522 ck: 2060 len: 152 ! Abg26522 Novel human diagnostic protein #26

(R,K) {20,20}

(R) {20}

46: EEEEE RRRRRRRRRRRRRRRRRR RKKE
(R) {20}

47: EEEER RRRRRRRRRRRRRRRRRR RKKEG
(R) {20}

48: EEEER RRRRRRRRRRRRRRRRRR KKEGE
(R) {20}

49: EEEER RRRRRRRRRRRRRRRRRR KEGEE
(R,K) {20}

50: EEEER RRRRRRRRRRRRRRRRRR KEGEG
(R,K) {20}

51: KKEEG RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

97: KEEGR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

98: EEEGR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

99: EEEGR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

100: GRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

101: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

102: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

103: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

104: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

105: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

106: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

107: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

108: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

109: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

110: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

111: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R) {20}

112: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R,K) {20}

113: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R,K) {20}

ABg26525 ck: 5433 len: 103 ! Abg26525 Novel human diagnostic protein #26

(R,K) {20,20}

(R) {20}

3: RD RRRKKKKKKKKKKKKKKKK EEEEE

ABg26526 ck: 1887 len: 115 ! Abg26526 Novel human diagnostic protein #26

(R,K) {20,20}

(R) {20}

48: EEEEE RRRRRRRRRRRRRRRRRR KRRK
(R) {20}

49: EEEER RRRRRRRRRRRRRRRRRR KRRK
(R,K) {20}

50: EEEER RRRRRRRRRRRRRRRRRR RRRK
(R,K) {20}

51: EEEER RRRRRRRRRRRRRRRRRR RRRK
(R,K) {20}

1
 ABG26527 ck: 5586 len: 122 i Abg26527 Novel human diagnostic protein #26
 (R,K){20,20}
 82: EEEEE KKKKKKKKKKKKKRRRRRR GRRRM
 1
 ABG26528 ck: 2237 len: 215 i Abg26528 Novel human diagnostic protein #26
 (R,K){20,20}
 87: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK
 (R,K){20}
 88: KEKEK KKKKKKKKKKKKKKKKKKK KKKKK
 (R,K){20}
 89: EKEKK RKKKKKKKKKKKKKKKKKK KKKKK
 (R,K){20}
 90: KEKKR KKKKKKKKKKKKKKKKKKK KKKKK
 (R,K){20}
 91: EKRRK KKKKKKKKKKKKKKKKKKK KKKKK
 (R,K){20}
 92: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (R,K){20}
 93: KKKKK RKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 94: RKKKK KKKKKKKKKKKKKKKKKKK KKEEE
 (K){20}
 95: KKKKK KKKKKKKKKKKKKKKKKKK KEEEE
 (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKK EEEEE
 (R){20}
 167: GRRRG RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 168: RRRGR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 169: RRRGR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 170: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 171: GRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 172: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 173: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 174: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 175: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 176: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}

1
 177: RRRRR RRRRRRRRRRRRRRRRRR RRKR
 ABG26530 ck: 5729 len: 404 i Abg26530 Novel human diagnostic protein #26
 (R,K){20,20}
 366: EEEEE RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 367: EEEER RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 368: EEEER RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 369: EEEER RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 370: ERRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 371: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 372: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 373: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 374: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 375: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 376: RRRRR RRRRRRRRRRRRRRRRRR RRKNT
 (R){20}
 377: RRRRR RRRRRRRRRRRRRRRRRR RKNTS
 (R){20}
 378: RRRRR RRRRRRRRRRRRRRRRRR KNTSL
 (R,K){20}
 379: RRRRR RRRRRRRRRRRRRRRRRR NTSLR
 ABG26531 ck: 7434 len: 126 i Abg26531 Novel human diagnostic protein #26
 (R,K){20,20}
 48: EEKDE RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 49: EKDER RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 50: KDERR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 51: DERRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 52: ERRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}
 54: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
 (R){20}

[illegible]

```

54: RTRRR RRRRRRRRRRRRRRRRRR RRRKK
      (R) {20}
55: TRRRR RRRRRRRRRRRRRRRRRR RKKKP
      (R) {20}
56: RRRRR RRRRRRRRRRRRRRRRRR KKKPT
      (R) {20}
57: RRRRR RRRRRRRRRRRRRRRRRR KKPTS
      (R, K) {20}
58: RRRRR RRRRRRRRRRRRRRRRRR KPTRR
      (R, K) {20}
59: RRRRR RRRRRRRRRRRRRRRRRR PSTRV
      (R, K) {20}

Abg26534 ch: 1919 len: 113 1 Abg26534 Novel human diagnostic protein #26
      (R, K) {20, 20}
31: KKKKE KKKKKRRKKKKKKKKKKKK KKKKK
      (R, K) {20}
32: KKKKK KKKRRKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
33: KKKKK RKKRRKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
34: KKKRR KKKRRKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
35: EKKKK KRRKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
36: KKKKK RRRKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
37: KKKRR RKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
38: RKKRR KRRKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
39: KKKRK RKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
40: KRRRR KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
41: RRRRK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
42: RRRKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
44: RKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}

```


61: EKETK RRRRRKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 62: KETKR RRRKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 63: ETGR RRRKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 64: TKRR RRRKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 65: KRRR RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 66: RRRR RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 67: RRRR RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 68: RRRK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 69: RRRK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 70: RKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 71: KKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K) {20}
 72: KKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}
 73: KKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}
 74: KKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}
 75: KKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}
 76: RKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}
 77: KKKK RKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

1
 ABG26539 ck: 3545 len: 89 ! Abg26539 Novel human diagnostic protein #26
 (R,K) {20,20}
 28: EEEEE RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 29: EEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 30: EEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 31: EEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 32: EEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 33: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}

34: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 35: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 36: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 37: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 38: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 39: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 40: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 41: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 42: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 43: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 44: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 45: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 46: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 47: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 48: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 49: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 50: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}

1
 ABG26542 ck: 4605 len: 182 ! Abg26542 Novel human diagnostic protein #26
 (R,K) {20,20}
 146: EGEE RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 147: GEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 148: EEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 149: EEEER RRRRRRRRRRRRRRRRRRRRRRRRR
 (R) {20}
 ABG26543 ck: 2998 len: 405 ! Abg26543 Novel human diagnostic protein #26
 (R,K) {20,20}
 264: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R,K) {20}

```

265: RRRWR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
266: RRWRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
267: RWRRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
268: WRRRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
269: RRRRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
270: RRRRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
271: RRRRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
272: RRRRR RRRRRRKKKKKKKKKKKK KKXKX      (R,K){20}
273: RRRRR RRRRRRKKKKKKKKKKKK KKXKX      (K){20}
274: RRRRK RRRRRRKKKKKKKKKKKK KKXKX      (K){20}
275: RRRKK RRRRRRKKKKKKKKKKKK KKXKX      (K){20}
276: RRRKK RRRRRRKKKKKKKKKKKK KKXKX      (K){20}
277: RRRKK RRRRRRKKKKKKKKKKKK KKXKX      (K){20}
278: RRRKK RRRRRRKKKKKKKKKKKK KKXKG      (K){20}
279: RRRKK RRRRRRKKKKKKKKKKKK KKXKK      (K){20}
280: RRRKK RRRRRRKKKKKKKKKKKK KKXKK      (K){20}
281: RRRKK RRRRRRKKKKKKKKKKKK KXKXG      (K){20}
282: RRRKK RRRRRRKKKKKKKKKKKK GKXKG      (K){20}

ABg26545 ck: 135 len: 119 ! Abg26545 Novel human diagnostic protein #26
          (R,K){20,20}
88: EREKE RRRRRRKKKKKKKKKKKK KKXKK      (K){20}
89: KEKEK RRRRRRKKKKKKKKKKKK KKXKK      (K){20}
90: EREKK RRRRRRKKKKKKKKKKKK KKXKK      (K){20}
91: KEKKK RRRRRRKKKKKKKKKKKK KKXKK      (K){20}
92: EKKKK RRRRRRKKKKKKKKKKKK KKXKK      (K){20}

```

```

93: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
94: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
95: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKNF
      (K) {20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKK KNF
      (K) {20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKK NF

ABG26551 ck: 8976 len: 94 | Abg26551 Novel human diagnostic protein #26
1
1:
      (R, K) {20, 20}
      (R, K) {20}
      KRRKKKKKKKKKKKKKKKKKKKK KKKKK
2: K RRRKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
3: KR RRRKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
4: KRR RRRKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
5: KRRR KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
6: KRRRK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
7: RRRKK KKKKKKKKKKKKKKKKKKKKK RKKKK
      (K) {20}
8: RRRKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
9: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK RKKKK
      (R, K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK RKKKK
      (R, K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R, K) {20}

```

18: KKKKK (R,K) {20} KKKKK
 19: KKKKK (R,K) {20} KKKKK
 20: KKKKK (R,K) {20} KKKKK
 21: KKKKK (R,K) {20} KKKKK

ABG26717 ck: 4415 len: 78 | Abg26717 Novel human diagnostic protein #26

37: OMLSV (R,K) {20,20} KKKKK
 38: MLSVK (R,K) {20} KKKKK
 39: LSVKK (R,K) {20} KKKKK
 40: SVKKK (R,K) {20} KKKKK
 41: VKKKK (R,K) {20} KKKKK
 42: KKKKK (R,K) {20} KKKKK
 43: KKKKK (R,K) {20} KKKKK
 44: KKKKK (R,K) {20} KKKKK
 45: KKKKK (R,K) {20} KKKKK
 46: KKKKK (R,K) {20} KKKKK
 47: RKKKK (R,K) {20} KKKKK
 48: KKKKK (R,K) {20} KKKKK
 49: KKKKK (R,K) {20} KKKKK
 50: KKKKK (R,K) {20} KKKKK
 51: KKKKK (R,K) {20} KKKKK
 52: KKKKK (R,K) {20} KKKKK
 53: KKKKK (R,K) {20} KKKKK
 54: KKKKK (R,K) {20} KKKKK

ABG26718 ck: 9531 len: 141 | Abg26718 Novel human diagnostic protein #26

80: EVARP (R,K) {20,20} KKKKK
 81: VAPRR (R,K) {20} KKKKK
 82: APRRK (R,K) {20} KKKKK
 83: RPRKK (R,K) {20} KKKKK
 84: PRKKK (R,K) {20} KKKKK
 85: RKKKK (R,K) {20} KKKKK
 86: KKKKK (R,K) {20} KKKKK
 87: KKKKK (R,K) {20} KKKKK
 88: KKKKK (R,K) {20} KKKKK
 89: KKKKK (R,K) {20} KKKKK

ABG26719 ck: 72 len: 83 | Abg26719 Novel human diagnostic protein #26

42: ETPSE (R,K) {20,20} KKKKK
 43: TPSEK (R,K) {20} KKKKK
 44: PSEKK (R,K) {20} KKKKK
 45: SEKKK (R,K) {20} KKKKK
 46: EKKKK (R,K) {20} KKKKK
 47: KKKKK (R,K) {20} KKKKK
 48: KKKKK (R,K) {20} KKKKK
 49: KKKKK (R,K) {20} KKKKK
 50: KKKKK (R,K) {20} KKKKK
 51: KKKKK (R,K) {20} KKKKK
 52: KKKKK (R,K) {20} KKKKK
 53: KKKKK (R,K) {20} KKKKK
 54: KKKKK (R,K) {20} KKKKK

1

ABG26720 ck: 3781 len: 57 ! Abg26720 Novel human diagnostic protein #26

(R,K){20,20}

(K){20}

24: QMKSI KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

25: MKSIK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

26: KSIKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

27: SIKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

28: IKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKF

(K){20}

30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKFH

(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKHF

(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKHFL

(K){20}

33: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKFHL

ABG26721 ck: 287 len: 95 ! Abg26721 Novel human diagnostic protein #26

(R,K){20,20}

(K){20}

31: EKEKE KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

32: KEKEK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

33: EKEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

34: KEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

35: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

1

ABG26722 ck: 4831 len: 127 ! Abg26722 Novel human diagnostic protein #26

(R,K){20,20}

(K){20}

44: KKKKE KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

45: KKEEK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

46: KEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKE

(K){20}

47: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKE

(K){20}

48: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKE

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKE

(K){20}

87: EKEEE KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

88: KEEEK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

89: EEEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

90: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

91: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
93: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
94: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
95: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
104: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
105: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
106: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
107: KKKKK KKKKKKKKKKKKKKKKKKKKK K
 (K) {20}
108: KKKKK KKKKKKKKKKKKKKKKKKKKK
 (K) {20}

ABG26723 ck: 7054 len: 98 ; Abg26723 Novel human diagnostic protein #26
 (R,K) {20,20}
42: EKEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
43: KEKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
44: EKEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
45: KEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
46: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

ABG26724 ck: 9897 len: 295 ; Abg26724 Novel human diagnostic protein #26

1

(R,K){20,20}
259: KEEBE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
260: EEEBEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
261: EEEBK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
262: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
263: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
264: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
265: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
266: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
267: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
268: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
269: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
270: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
271: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
272: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
273: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
274: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
275: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
276: KKKKK KKKKKKKKKKKKKKKKKKKKK
ABG26725 ck: 909 len: 131 ! Abg26725 Novel human diagnostic protein #26
(R,K){20,20}
70: KKKKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1

74: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
81: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
82: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
ABG26726 ck: 9300 len: 677 ! Abg26726 Novel human diagnostic protein #26
(R,K){20,20}
240: EKEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
241: KEKKR KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
242: EKEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
243: KEKKR KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
244: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
245: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
246: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
247: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
248: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
249: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
250: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
251: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
252: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1

253: RKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
254: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
255: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
256: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
257: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
258: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
259: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
260: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
261: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
262: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
263: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
264: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
265: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
ABG26727 ck: 813 len: 329 ! Abg26727 Novel human diagnostic protein #26
(R,K) {20,20}
215: EKEKE KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
216: KEKEK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
217: EKEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
218: KEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
219: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
220: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
221: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
222: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
223: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
224: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

1

1

1

225: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
226: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
227: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
228: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
229: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
230: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
ABG26879 ck: 2669 len: 800 ! Abg26879 Novel human diagnostic protein #26
(R,K) {20,20}
559: EREGE KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
560: REOEK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
561: EOEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
562: OEKKR KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
563: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
564: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
ABG28885 ck: 8668 len: 137 ! Abg28885 Novel human diagnostic protein #28
(R,K) {20,20}
35: EEEEG RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R) {20}
36: EEEGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R) {20}
37: EEEGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R) {20}
38: EGGRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R) {20}
39: GRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R) {20}
40: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R,K) {20}
41: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R,K) {20}
AAG65985 ck: 8085 len: 154 ! Aag65985 B726P splice variant sequence. 2/2
(R,K) {20,20}
114: TQLRQ KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
115: QLROK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

1

116: LROKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

117: ROKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

118: OKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

119: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

120: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

121: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

122: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

123: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

124: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

125: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

126: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

127: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

128: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

ABB27893 ck: 5383 len: 86 ! Abb27893 Human peptide #544 encoded by brea

15: RRRRG RRRRRKKKKKKKKKKKKKK (R, K) {20, 20}

16: RRRGR RRRKKKKKKKKKKKKKKKK (R, K) {20}

17: RRGRR RRRKKKKKKKKKKKKKKKK (R, K) {20}

18: RGRRR RRRKKKKKKKKKKKKKKKK (R, K) {20}

19: GRRRR RKKKKKKKKKKKKKKKKKK (R, K) {20}

20: RRRRR KKKKKKKKKKKKKKKKKRR (R, K) {20}

21: RRRRK KKKKKKKKKKKKKKKKKRRRR (R, K) {20}

22: RRRKK KKKKKKKKKKKKKKKRRRRR (R, K) {20}

23: RRRKK KKKKKKKKKKKKKRRRRRR (R, K) {20}

24: RKKKK KRRKKKKKKKKRRRRRRRR (R, K) {20}

1

25: KKKKK KKKKKKKKKKKRRRRRRRR (R, K) {20}

26: KKKKK RKKKKKKKKRRRRRRRRRR (R, K) {20}

27: KKKKK KKKKKKKKKRRRRRRRRRR (R, K) {20}

28: KKKKK KKKKKKKKKRRRRRRRRRR (R, K) {20}

29: KKKKK KKKKKKKKKRRRRRRRRRR (R, K) {20}

30: KKKKK KKKKKKKKKRRRRRRRRRR (R, K) {20}

31: RKKKK KKKKKKKKKRRRRRRRRRR (R, K) {20}

32: KKKKK RKKKKRRRRRRRRRRRRRR (R, K) {20}

33: KKKKK KKKKKRRRRRRRRRRRRRR (R, K) {20}

34: KKKKK KKKRRRRRRRRRRRRRRRR (R, K) {20}

35: KKKKK KKKRRRRRRRRRRRRRRRR (R, K) {20}

36: KKKKK KRRRRRRRRRRRRRRRRRR (R, K) {20}

37: RKKKK RRRRRRRRRRRRRRRRRRR (R) {20}

38: KKKKK RRRRRRRRRRRRRRRRRRR (R) {20}

39: KKKKK RRRRRRRRRRRRRRRRRRR (R) {20}

40: KKKRR RRRRRRRRRRRRRRRRRRR (R) {20}

41: KRRRR RRRRRRRRRRRRRRRRRRR (R) {20}

42: RRRRR RRRRRRRRRRRRRRRRRRR (R) {20}

43: RRRRR RRRRRRRRRRRRRRRRRRR (R) {20}

44: RRRRR RRRRRRRRRRRRRRRRRRR (R) {20}

45: RRRRR RRRRRRRRRRRRRRRRRRR (R) {20}

ABB28750 ck: 1334 len: 86 ! Abb28750 Peptide #1401 encoded by breast ce

57: EEEEG RRRKKKKKKKKKKKKKKKK (R, K) {20, 20}

58: EEEGR RRRKKKKKKKKKKKKKKKK (R, K) {20}

59: EEGRR $(R,K)\{20\}$ RKKKKRKKKKKKKKKKKK KKKK
60: EGRRR $(R,K)\{20\}$ KKKKKRKKKKKKKKKKKK KKKK
61: GRRRR $(R,K)\{20\}$ KKKKKRKKKKKKKKKKKK KKKK
62: RRRKK $(R,K)\{20\}$ KKKKKRKKKKKKKKKKKK KKKK
63: RRRKK $(R,K)\{20\}$ KKKKKRKKKKKKKKKKKK KKKK
64: RKKKK $(R,K)\{20\}$ RKKKKKKKKKKKKKKKK KKK
65: KKKKR $(R,K)\{20\}$ KKKKKKKKKKKKKKKKK KK
66: KKKKK $(R,K)\{20\}$ RKKKKKKKKKKKKKKKK K
67: KKKRR $(K)\{20\}$ KKKKKKKKKKKKKKKKK
1
ABB28840 ck: 9082 len: 167 1 Abb28840 Peptide #1491 encoded by breast ce
33: EEGRG $(R,K)\{20,20\}$ RRRRRRRRRRRRRRRRR RRGCG
34: EGRGR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RGGGR
35: GRRGR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR GGGRR
1
ABB29360 ck: 2276 len: 89 1 Abb29360 Peptide #2011 encoded by breast ce
23: EEEEE $(R,K)\{20,20\}$ KKKKKRKKKKKKKKKK EEEKK
46: KKEEE $(K)\{20\}$ KKKKKKKKKKKKKKKKK KRKEE
47: KEEEE $(K)\{20\}$ KKKKKKKKKKKKKKKKK RKEEE
48: EEEKK $(R,K)\{20\}$ KKKKKKKKKKKKKKKKK KEEEE
49: EEEKK $(R,K)\{20\}$ KKKKKKKKKKKKKKKKK EEEEE
1
ABB29645 ck: 1939 len: 130 1 Abb29645 Peptide #2296 encoded by breast ce
42: EGRRK $(R,K)\{20,20\}$ RRRRRRRRRRRRRRRRR RRRRR
43: GRKKR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RRRRR
44: RKKRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RRRRR

45: KERRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RRRKK
46: EERRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RRRKK
47: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RKKKK
48: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RKKKK
49: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RKKKK
50: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RKKKE
51: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RKKKE
52: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR RKKKE
53: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR KEEEE
54: RRRRR $(R,K)\{20\}$ RRRRRRRRRRRRRRRRR EEEEE
1
ABB29928 ck: 3607 len: 88 1 Abb29928 Peptide #2579 encoded by breast ce
39: ERKRE $(R,K)\{20,20\}$ KKKKKKKKKKKKKKKKK KKKKK
40: RRRKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
41: KREKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
42: REKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
43: EKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
44: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
45: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
46: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
47: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
48: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
49: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
50: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK
51: KKKKK $(K)\{20\}$ KKKKKKKKKKKKKKKKK KKKKK

52: KKKK (K) {20}
53: KKKK (K) {20}
54: KKKK (K) {20}
55: KKKK (K) {20}
56: KKKK (K) {20}
57: KKKK (K) {20}
58: KKKK (K) {20}
59: KKKK (K) {20}
60: KKKK (K) {20}
61: KKKK (K) {20}
62: KKKK (K) {20}
63: KKKK (K) {20}
64: KKKK (K) {20}
65: KKKK (K) {20}
66: KKKK (K) {20}
67: KKKK (K) {20}
68: KKKK (K) {20}
69: KKKK (K) {20}

ABb30512 ck: 3937 len: 85 ! Abb30512 Peptide #3163 encoded by breast ce

1

1: (R,K) {20,20}
2: K (K) {20}
3: K (K) {20}
4: K (K) {20}
5: K (K) {20}

6: KKKK (K) {20}
7: KKKK (K) {20}
8: KKKK (K) {20}
9: KKKK (K) {20}
10: KKKK (K) {20}
11: KKKK (K) {20}

ABb32308 ck: 1560 len: 88 ! Abb32308 Peptide #4959 encoded by breast ce

1

43: RRRG (R,K) {20,20}
44: RRRG (R) {20}
45: RRRG (R) {20}
46: RRRG (R) {20}
47: RRRG (R) {20}
48: RRRG (R) {20}
49: RRRG (R) {20}
50: RRRG (R) {20}
51: RRRG (R) {20}
52: RRRG (R) {20}
53: RRRG (R) {20}
54: RRRG (R) {20}
55: RRRG (R) {20}
56: RRRG (R) {20}
57: RRRG (R) {20}
58: RRRG (R) {20}
59: RRRG (R) {20}

[illegible]

35: KKKKK KKKRRRRRRRRRRRRRRRRRR (R,K){20}
36: KKKKK KRRRRRRRRRRRRRRRRRR (R,K){20}
37: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
38: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
39: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
40: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
41: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
42: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
43: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
44: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}
45: KKKKK RRRRRRRRRRRRRRRRRRR (R){20}

1
ABB33937 ck: 1334 len: 86 | Abb33937 Peptide #1443 encoded by human foet

57: EEEEG RRRRRRRRRRRRRRRRRRR (R,K){20,20}
58: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
59: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
60: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
61: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
62: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
63: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
64: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
65: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
66: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}
67: EEEGR RRRRRRRRRRRRRRRRRRR (R,K){20}

ABB34024 ck: 9082 len: 167 | Abb34024 Peptide #1530 encoded by human foet

1

(R,K){20,20}
33: EEEGRG RRRRRRRRRRRRRRRRRR RRGGR
(R,K){20}
34: EGRGR RRRRRRRRRRRRRRRRRR RGGGR
(R,K){20}
35: GGRGR RRRRRRRRRRRRRRRRRR GGGR

1

ABB34533 ck: 2276 len: 89 1 Abb34533 Peptide #2039 encoded by human foe
(R,K){20,20}
23: EEEEE KKKKKKKKKKKKKKKKKKK EEEK
(R,K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKK KRKEE
(K){20}
47: KEEER KKKKKKKKKKKKKKKKKKK RKEEE
(R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKK KEEEE
(R,K){20}
49: EEEKK KKKKKKKKKKKKKKKKKKK EEEEE

1

ABB34819 ck: 1939 len: 130 1 Abb34819 Peptide #2325 encoded by human foe
(R,K){20,20}
42: EGRKE RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
43: GRKER RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
44: RKEER RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
45: KEERR RRRRRRRRRRRRRRRRRR RREKK
(R,K){20}
46: ERRER RRRRRRRRRRRRRRRRRR RRRKK
(R,K){20}
47: RRRER RRRRRRRRRRRRRRRRRR RKKKK
(R,K){20}
48: RRRER RRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
49: RRRER RRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}
50: RRRER RRRRRRRRRRRRRRRRRR KKKKE
(R,K){20}
51: RRRER RRRRRRRRRRRRRRRRRR KKKKE
(R,K){20}
52: RRRER RRRRRRRRRRRRRRRRRR KKEEE
(R,K){20}
53: RRRER RRRRRRRRRRRRRRRRRR KEEEE
(R,K){20}
54: RRRER RRRRRRRRRRRRRRRRRR EEEEE

1

ABB35110 ck: 3607 len: 88 1 Abb35110 Peptide #2616 encoded by human foe
(R,K){20,20}
39: ERKRE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
40: RKREK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
41: KREKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: REKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K) {20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK

ABB35676 ck: 3937 len: 85 ! Abb35676 Peptide #3182 encoded by human foe

1: (R,K) {20,20}
(K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK EEEEX
(K) {20}

ABB36406 ck: 2686 len: 71 ! Abb36406 Peptide #3912 encoded by human foe

(R,K) {20,20}
(K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

23: KKKKK (K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK (K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
25: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
26: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
27: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
28: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
29: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
30: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
31: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
32: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
33: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
34: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
35: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
36: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
37: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
38: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
39: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
40: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
41: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
42: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
43: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
44: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
45: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
46: KKKKK (R,K) {20}
K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

47: KKKKK KKKKKKKKKKKKKKKKK KKSAN
(K) {20}
48: KKKKK KKKKKKKKKKKKKKKKK KSAH
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKK SAH
(K) {20}

ABB37567 ck: 1560 len: 88 1 Abb37567 Peptide #5073 encoded by human foe
(R,K) {20,20}

43: RRRRG RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
44: ERRGR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
45: RRGRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
46: RGRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
47: GRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
48: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
49: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
50: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
51: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
52: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
53: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
54: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
55: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
56: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
57: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
58: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
59: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
60: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
61: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
62: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}

1
ABB37780 ck: 2324 len: 36 1 Abb37780 Peptide #5286 encoded by human foe
(R,K) {20,20}

16: KERKT KKKRRRRRRRRRRRRRR R
(R,K) {20}
17: ERKTX KKKRRRRRRRRRRRRRR R
(R,K) {20}

1
ABB40272 ck: 8343 len: 66 1 Abb40272 Peptide #7778 encoded by human foe
(R,K) {20,20}

6: ETERE KKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
7: TEREK KKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
8: EREKK KKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
9: REKKK KKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}

1
ABB42642 ck: 2394 len: 57 1 Abb42642 Peptide #10148 encoded by human foe
(R,K) {20,20}

20: EEEEG RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
21: EEEGR RRRRRRRRRRRRRRRRR RRRRG
(R) {20}
22: EEEGR RRRRRRRRRRRRRRRRR RRRGR
(R) {20}
23: EGGRR RRRRRRRRRRRRRRRRR RRRGR
(R) {20}
24: GRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}
25: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R) {20}

1
ABB43181 ck: 4228 len: 24 1 Abb43181 Peptide #10687 encoded by human foe
(R,K) {20,20}

1: RRRRRRRRRRRRRRRRRRR RRRRR
(R,K) {20}
2: RRRRRRRRRRRRRRRRRRR RRRRR
(R,K) {20}
3: RRRRRRRRRRRRRRRRRRR RRRRR
(R,K) {20}
4: RRRRRRRRRRRRRRRRRRR RRRRR
(R,K) {20}

1
ABB44317 ck: 4695 len: 51 1 Abb44317 Peptide #11823 encoded by human foe
(R,K) {20,20}

18: LFKPM KKKRRRRRRRRRRRRRR KKLTT
(R,K) {20}
19: FKPMR KKKRRRRRRRRRRRRRR KKLTT
(R,K) {20}

20: KPMRK RRRKKRRKKKKRRRRKK LTTT
(R,K){20}

ABBI7165 ck: 8887 len: 42 I Abbi7165 Human nervous system related poly

(R,K){20,20}

21: FTTTE KKKKKKKKKKKKKKKKKKK KX
(K){20}

22: FTTEK KKKKKKKKKKKKKKKKKKK X
(K){20}

ABBI8534 ck: 5383 len: 86 I Abbi8534 Protein #533 encoded by probe for

(R,K){20,20}

15: RRRRG RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}

16: RRRGR RRRKKKKKKKKKKKKKKKK KRRR
(R,K){20}

17: RRGRR RRRKKKKKKKKKKKKKKKK RRRR
(R,K){20}

18: RGRRR RRRKKKKKKKKKKKKKKKK RRRR
(R,K){20}

19: GRRRR RKKKKKKKKKKKKKKKKKK RRRR
(R,K){20}

20: RRRRR KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

21: RRRRK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

22: RRRKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

23: RRRKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

24: RKKKK KRRKKKKKKKKKKKKKKRR RRRR
(R,K){20}

25: KKKKK KRRKKKKKKKKKKKKKKRR RRRR
(R,K){20}

26: KKKKK RKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

27: KKKKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

28: KKKKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

29: KKKKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

30: KKKKK KRRKKKKKKKKKKKKKKRR RRRR
(R,K){20}

31: RKKKK KRRKKKKKKKKKKKKKKRR RRRR
(R,K){20}

32: KKKKK RKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

33: KKKKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

34: KKKKK KKKKKKKKKKKKKKKKKRR RRRR
(R,K){20}

35: KKKKK KRRKKKKKKKKKKKKKKRR RRRR
(R,K){20}

36: KKKKK KRRKKKKKKKKKKKKKKRR RRRR
(R,K){20}

37: RKKKK RRRRRRRRRRRRRRRRRR RRRR
(R){20}

38: KKKKK RRRRRRRRRRRRRRRRRR RRRR
(R){20}

39: KKKKK RRRRRRRRRRRRRRRRRR RRRR
(R){20}

40: KKKKK RRRRRRRRRRRRRRRRRR RRRR
(R){20}

41: KRRRR RRRRRRRRRRRRRRRRRR RRRR
(R){20}

42: RRRRR RRRRRRRRRRRRRRRRRR RRRR
(R){20}

43: RRRRR RRRRRRRRRRRRRRRRRR RRRR
(R){20}

44: RRRRR RRRRRRRRRRRRRRRRRR RRRR
(R){20}

45: RRRRR RRRRRRRRRRRRRRRRRR RRRR
(R){20}

ABBI9373 ck: 1334 len: 86 I Abbi9373 Protein #1372 encoded by probe for

(R,K){20,20}

57: EEEEG RRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

58: EEEGR RRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

59: EEEGR RRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

60: EEEGR KRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

61: GRRRK KRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

62: RRRKK KRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

63: RRRKK KRRKKKKKKKKKKKKKKKK KKKK
(R,K){20}

64: RKKKK RRRKKKKKKKKKKKKKKKK KKK
(R,K){20}

65: KKKKK KRRKKKKKKKKKKKKKKKK KK
(R,K){20}

66: KKKKK RRRKKKKKKKKKKKKKKKK K
(R,K){20}

67: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}

1
 ABB19467 ck: 9082 len: 167 i Abb19467 Protein #1466 encoded by probe for
 (R,K){20,20}
 33: EEEGRG RRRRRRRRRRRRRRRRRRRR RRGCG
 (R,K){20}
 34: EGRGR RRRRRRRRRRRRRRRRRRRR RGGGR
 (R,K){20}
 35: GGRGR RRRRRRRRRRRRRRRRRRRR GGGRR
 (R,K){20}
 ABB1943 ck: 2276 len: 89 i Abb1943 Protein #1942 encoded by probe for
 (R,K){20,20}
 23: EEEEE KKKKKKKKKKKKKKKKKKK EEEKK
 (R,K){20}
 46: KKEEE KKKKKKKKKKKKKKKKKKK KRKEE
 (K){20}
 47: KEEKK KKKKKKKKKKKKKKKKKKK RKEEE
 (K){20}
 48: EEEKK KKKKKKKKKKKKKKKKKKK KEEEE
 (R,K){20}
 49: EEEKK KKKKKKKKKKKKKKKKKKK EEEEE
 (R,K){20}
 ABB20238 ck: 1939 len: 130 i Abb20238 Protein #2237 encoded by probe for
 (R,K){20,20}
 42: EGRKE RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 43: GRKER RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 44: RKERR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 45: KEERR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 46: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
 53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R,K){20}

1
 ABB20531 ck: 3607 len: 88 i Abb20531 Protein #2530 encoded by probe for
 (R,K){20,20}
 39: ERKRE KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 40: RKREK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 41: KREKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 42: REKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 43: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

61: KKKK (K) {20} KKKK
 62: KKKK (K) {20} KKKK
 63: KKKK (K) {20} KKKK
 64: KKKK (K) {20} KKKK
 65: KKKK (K) {20} KKKK
 66: KKKK (K) {20} KKKK
 67: KKKK (K) {20} KKKK
 68: KKKK (K) {20} K
 69: KKKK (K) {20} KKKK

ABb21105 ck: 3937 len: 85 ! Abb21105 Protein #3104 encoded by probe for

1

1: KKKK (R, K) {20,20} KKKK
 2: K KKKK (K) {20} KKKK
 3: KK KKKK (K) {20} KKKK
 4: KK KKKK (K) {20} KKKK
 5: KKKK (K) {20} KKKK
 6: KKKK (K) {20} KKKK
 7: KKKK (K) {20} KKKK
 8: KKKK (K) {20} KKKK
 9: KKKK (K) {20} KKKK
 10: KKKK (K) {20} KKKK
 11: KKKK (K) {20} EEEEX

ABb21763 ck: 2686 len: 71 ! Abb21763 Protein #3762 encoded by probe for

1

20: KKKKQ KKKK (R, K) {20,20} KKKK
 21: KKKK (K) {20} KKKK

22: KKQK (K) {20} KKKK
 23: KQKK (K) {20} KKKK
 24: QKKK (K) {20} KKKK
 25: KKKK (R, K) {20} KKKK
 26: KKKK (R, K) {20} KKKK
 27: KKKK (R, K) {20} KKKK
 28: KKKK (R, K) {20} KKKK
 29: KKKK (R, K) {20} KKKK
 30: KKKK (R, K) {20} KKKK
 31: KKKK (R, K) {20} KKKK
 32: KKKK (R, K) {20} KKKK
 33: KKKK (R, K) {20} KKKK
 34: KKKK (R, K) {20} KKKK
 35: KKKK (R, K) {20} KKKK
 36: KKKK (R, K) {20} KKKK
 37: KKKK (R, K) {20} KKKK
 38: KKKK (R, K) {20} KKKK
 39: KKKK (R, K) {20} KKKK
 40: KKKK (R, K) {20} KKKK
 41: KKKK (R, K) {20} KKKK
 42: KKKK (R, K) {20} KKKK
 43: KKKK (R, K) {20} KKKK
 44: KKKK (R, K) {20} KKKK
 45: KKKK (R, K) {20} KKKK

46: KKKK (R,K){20} KKKK
 47: KKKR (K){20} KKSAA
 48: KKKK (K){20} KSAH
 49: KKKK (K){20} SAH

1 ABB22862 ck: 1560 len: 88 ! ABB22862 Protein #4661 encoded by probe for
 (R,K){20,20}

43: RRRG (R,K){20} RRRR
 44: ERGR (R){20} RRRR
 45: RGRG (R){20} RRRR
 46: RGRG (R){20} RRRR
 47: GRRR (R){20} RRRR
 48: RRRR (R){20} RRRR
 49: RRRR (R){20} RRRR
 50: RRRR (R){20} RRRR
 51: RRRR (R){20} RRRR
 52: RRRR (R){20} RRRR
 53: RRRR (R){20} RRRR
 54: RRRR (R){20} RRRR
 55: RRRR (R){20} RRRR
 56: RRRR (R){20} RRRR
 57: RRRR (R){20} RRRR
 58: RRRR (R){20} RRRR
 59: RRRR (R){20} RRRR
 60: RRRR (R){20} RRRR
 61: RRRR (R){20} RRRR

62: RRRR (R){20} NTNNE
 ABB23064 ck: 2324 len: 36 ! ABB23064 Protein #5063 encoded by probe for
 (R,K){20,20}
 16: KERK (R,K){20} R
 17: ERK (R,K){20} RRRR

1 ABB24685 ck: 8343 len: 66 ! ABB24685 Protein #6684 encoded by probe for
 (R,K){20,20}

6: ETERE (R,K){20} KKKK
 7: TEREK (R,K){20} KKKK
 8: EREK (R,K){20} KKKK
 9: REKK (R,K){20} KKKK

1 ABB25988 ck: 2394 len: 57 ! ABB25988 Protein #7987 encoded by probe for
 (R,K){20,20}

20: EEEG (R,K){20} RRRR
 21: EEEG (R){20} RRRR
 22: EGRG (R){20} RRRR
 23: EGRR (R){20} RRRR
 24: GRRR (R){20} RRRR
 25: RRRR (R){20} RRRR

1 ABB27176 ck: 4895 len: 51 ! ABB27176 Protein #9175 encoded by probe for
 (R,K){20,20}

18: LFKP (R,K){20} KLLT
 19: FKPM (R,K){20} KLLT
 20: KPMK (R,K){20} LTTT

1 ABB10296 ck: 3983 len: 292 ! ABB10296 Human cDNA SEQ ID NO: 604. 1/2002
 (R,K){20,20}

273: QVAP (R,K){20} KKKK

ABB10485 ck: 7611 len: 315 ! ABB10485 Human cDNA SEQ ID NO: 793. 1/2002

1

(R,K){20,20}
(R,K){20}
273: QVFP RKXXXXXXXXXXXXXXXXX KGRS
(K){20}
274: VFAPR KXXXXXXXXXXXXXXXXX GGRSR

AAU21948 ck: 444 len: 66 i Aau21948 Human cardiovascular system antige

1

(R,K){20,20}
(K){20}
35: SMTFS KXXXXXXXXXXXXXXXXX KXGKX
(K){20}
36: MTFSK KXXXXXXXXXXXXXXXXX XGKXK

AAU22148 ck: 4665 len: 34 i Aau22148 Human cardiovascular system antige

1

(R,K){20,20}
(K){20}
10: PELLK KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
11: ELLK KXXXXXXXXXXXXXXXXX KKKK
(K){20}
12: LLLK KXXXXXXXXXXXXXXXXX KKK
(K){20}
13: LLKK KXXXXXXXXXXXXXXXXX KK
(K){20}
14: LKKK KXXXXXXXXXXXXXXXXX K
(K){20}
15: KKKK KXXXXXXXXXXXXXXXXX

AAU22186 ck: 269 len: 76 i Aau22186 Human cardiovascular system antige

1

(R,K){20,20}
(K){20}
37: TPSRA KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
38: PSRAK KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
39: SRAK KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
40: RAKK KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
41: AKKK KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
42: KKKK KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
43: KKKK KXXXXXXXXXXXXXXXXX KKKKI
(K){20}
44: KKKK KXXXXXXXXXXXXXXXXX XKIK

AAU22374 ck: 8278 len: 53 i Aau22374 Human cardiovascular system antige

1

(R,K){20,20}
(K){20}

30: NCGIL KXXXXXXXXXXXXXXXXX KKKK

(K){20}

31: CGILK KXXXXXXXXXXXXXXXXX KKK

(K){20}

32: GILKK KXXXXXXXXXXXXXXXXX KK

(K){20}

33: ILKKK KXXXXXXXXXXXXXXXXX K

(K){20}

34: LKKK KXXXXXXXXXXXXXXXXX

AAU23799 ck: 6158 len: 272 i Aau23799 Novel human enzyme polypeptide #885

1

(R,K){20,20}
(K){20}
238: SPANA KXXXXXXXXXXXXXXXXX KKKKG
(K){20}
239: PANAK KXXXXXXXXXXXXXXXXX KKKGR
(K){20}
240: ANAKK KXXXXXXXXXXXXXXXXX KKGRR
(K){20}
241: NAKKK KXXXXXXXXXXXXXXXXX KGRPX
(K){20}
242: AKKK KXXXXXXXXXXXXXXXXX GRPXX

AAU27944 ck: 1121 len: 69 i Aau27944 Human config polypeptide sequence

1

(R,K){20,20}
(R,K){20}
40: VPPLT KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
41: PPLTR KXXXXXXXXXXXXXXXXX KKKKX
(K){20}
42: PLTRK KXXXXXXXXXXXXXXXXX KKKKR
(K){20}
43: LTRKK KXXXXXXXXXXXXXXXXX KKKRG
(K){20}
44: TRKKK KXXXXXXXXXXXXXXXXX KKRGA
(K){20}
45: RKKKK KXXXXXXXXXXXXXXXXX KRGAL
(K){20}
46: KKKKK KXXXXXXXXXXXXXXXXX RGAL
(R,K){20}
47: KKKKK KXXXXXXXXXXXXXXXXX GAL

AAU31467 ck: 4264 len: 657 i Aau31467 Novel human secreted protein #1958

1

(R,K){20,20}
(R){20}
19: RRRRP RRRRRRRRRRRRRRRRRR RRRRL
(R){20}
20: RRRPR RRRRRRRRRRRRRRRRRR RRRLG
(R){20}
21: RRRPR RRRRRRRRRRRRRRRRRR RRLGL

22: RPRRR RRRRRRRRRRRRRRRR RLGLE (R) {20}
23: PRRRR RRRRRRRRRRRRRRRR LGLE (R) {20}

AAU3348 ck: 8085 len: 154 ! AAU3348 Human breast cancer protein encode
(R,K) {20,20}

114: TQLRQ KKKKKKKKKKKKKKKKK (K) {20}

115: QLRRQ KKKKKKKKKKKKKKKKK (K) {20}

116: LRQKK KKKKKKKKKKKKKKKKK (K) {20}

117: RQKKK KKKKKKKKKKKKKKKKK (K) {20}

118: QKKKK KKKKKKKKKKKKKKKKK (K) {20}

119: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

120: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

121: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

122: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

123: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

124: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

125: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

126: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

127: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

128: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

AAE09664 ck: 1663 len: 87 ! Aae09664 Human pancreatic related protein H
(R,K) {20,20}

36: KMSXX KKKKKKKKKKKKKKKKK (K) {20}

37: WSSXX KKKKKKKKKKKKKKKKK (K) {20}

38: SSXXK KKKKKKKKKKKKKKKKK (K) {20}

39: SXKKK KKKKKKKKKKKKKKKKK (K) {20}

40: XXXXX KKKKKKKKKKKKKKKKK (K) {20}

41: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

42: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

43: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

44: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

45: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

46: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

47: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

48: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

49: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

50: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

51: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

52: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

53: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

54: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

55: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

56: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

57: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

58: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

59: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

60: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

61: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

62: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

63: KKKKK KKKKKKKKKKKKKKKKK (K) {20}

AAW95365 ck: 5626 len: 139 ! Aaw95365 Human reproductive system related

```

1      (R,K){20,20}
      (K){20}
111: IHLNL KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
112: HLNLK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
113: LNLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
114: NLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAM96607 ck: 4751 len: 80  ! Aam96607 Human reproductive system related
      (R,K){20,20}
      (K){20}
61: KKKFD KKKKKKKKKKKKKKKKKKKKK

AAU18162 ck: 7907 len: 39  ! Aau18162 Novel human DNA-binding protein #9
      (R,K){20,20}
      (K){20}
9: YFEDL KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
10: FEDLK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
11: EDLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
12: DLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
13: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18167 ck: 9194 len: 87  ! Aau18167 Novel human DNA-binding protein #1
      (R,K){20,20}
      (K){20}
52: KIILL KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
53: IILLK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
54: ILLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
55: LLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
56: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18168 ck: 8659 len: 104  ! Aau18168 Novel human DNA-binding protein #1
      (R,K){20,20}
      (K){20}
75: PLGGQ KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

```

```

76: LGGQK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
77: GGQKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
78: GQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
79: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18171 ck: 9398 len: 48  ! Aau18171 Novel human DNA-binding protein #1
      (R,K){20,20}
      (K){20}
2: Q KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
3: QK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
4: QKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18177 ck: 8278 len: 53  ! Aau18177 Novel human DNA-binding protein #2
      (R,K){20,20}
      (K){20}
30: NCGIL KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
31: CGILK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
32: GILKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
33: ILKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18178 ck: 444 len: 66  ! Aau18178 Novel human DNA-binding protein #2
      (R,K){20,20}
      (K){20}
35: SMTFS KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
36: MTFSS KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18179 ck: 5503 len: 50  ! Aau18179 Novel human DNA-binding protein #2
      (R,K){20,20}
      (K){20}
30: IICLL KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
31: ICLLK KKKKKKKKKKKKKKKKKKKKK KKKKK

AAU18184 ck: 5691 len: 108  ! Aau18184 Novel human DNA-binding protein #3
      (R,K){20,20}
      (K){20}
78: VRPCL KKKKKKKKKKKKKKKKKKKKK KKKKK

```

79: RPCLK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 80: PCLK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 81: CLKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 82: LKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 83: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 84: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 85: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 86: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 87: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 88: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 89: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18192 ck: 6029 len: 63 i Aau18192 Novel human DNA-binding protein #3
 (R, K) {20,20}
 (K) {20}
 40: KLTLK KKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18200 ck: 7170 len: 63 i Aau18200 Novel human DNA-binding protein #4
 (R, K) {20,20}
 (K) {20}
 37: TPSRA KKKKKKKKKKKKKKKKKKKKK (K) {20}
 38: PSRAK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 39: SRAKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 40: RAKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 41: AKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 42: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 43: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 44: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18204 ck: 6110 len: 61 i Aau18204 Novel human DNA-binding protein #5
 (R, K) {20,20}
 (K) {20}

28: RPTRP KKKKKKKKKKKKKKKKKKKKK (K) {20}
 29: PTPRP KKKKKKKKKKKKKKKKKKKKK (K) {20}
 30: TRPKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 31: RPKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 32: PKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 33: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 34: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 35: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 36: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 37: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18205 ck: 5764 len: 74 i Aau18205 Novel human DNA-binding protein #5
 (R, K) {20,20}
 (K) {20}

40: EFLSA KKKKKKKKKKKKKKKKKKKKK (K) {20}
 41: FLSAK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 42: LSARK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 43: SAKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 44: AKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 45: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 48: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 49: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 50: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 51: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKKK (K) {20}

53: KKKKK KKKKKKKKKKKKKKKKKKKKK KK

54: KKKKK (K) {20}
KKKKKKKKKKKKKKKKKKKK X

AAU18206 ck: 9217 len: 68 1 Aau18206 Novel human DNA-binding protein #5

(R,K) {20,20}

38: FLFPE KKKKKKKKKKKKKKKKKKKKK KKKKK

39: LPPEK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

40: FPEKK KKKKKKKKKKKKKKKKKKKKK KKKKK

41: PEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

42: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

45: KKKKK KKKKKKKKKKKKKKKKKKKKK GKXX

AAU18208 ck: 8162 len: 79 1 Aau18208 Novel human DNA-binding protein #5

(R,K) {20,20}

41: VRPRV RKKKKKKKKKKKKKKKKKKKK KKKKK

42: RPRVR KKKKKKKKKKKKKKKKKKKKK KKKKK

43: PRVRK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

44: RVRKK KKKKKKKKKKKKKKKKKKKKK KKKKK

45: VRKKK KKKKKKKKKKKKKKKKKKKKK GGRFR

AAU18237 ck: 285 len: 118 1 Aau18237 Novel human DNA-binding protein #8

(R,K) {20,20}

98: EKHKQ KKKKKKKKKKKKKKKKKKKKK R G

AAU18238 ck: 5509 len: 58 1 Aau18238 Novel human DNA-binding protein #8

(R,K) {20,20}

36: FYFVC KKKKKKKKKKKKKKKKKKKKK KKK

37: YFVCK KKKKKKKKKKKKKKKKKKKKK KK

38: FVCKK KKKKKKKKKKKKKKKKKKKKK K

(K) {20}

39: VCKKK KKKKKKKKKKKKKKKKKKKKK

AAU18239 ck: 9074 len: 66 1 Aau18239 Novel human DNA-binding protein #8

(R,K) {20,20}

40: LVQCE KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

41: VQCEK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

42: QCEKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

43: CEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

44: EKKKK KKKKKKKKKKKKKKKKKKKKK KKK

(K) {20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKK KK

(K) {20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKK K

(K) {20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKK

AAU18240 ck: 8528 len: 150 1 Aau18240 Novel human DNA-binding protein #8

(R,K) {20,20}

113: SRNTV KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

114: RNTVK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

115: NTVRK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

116: TVRKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

117: VRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

118: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

119: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

120: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

121: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

122: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

123: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

AAU18241 ck: 7676 len: 156 1 Aau18241 Novel human DNA-binding protein #8

(R,K) {20,20}

108: KTTWI KKKKKKKKKKKKKKKKKKKKK KKKKK

109: TWIWK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
110: TWIWK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
111: WIKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKRG
112: IKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKRGX
113: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KRGXG
114: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} RKGXS
115: KKKKK KKKKKKKKKKKKKKKKKKK (R,K) {20} GKGST

1
AAU18242 ck: 1736 len: 40 1 Aau18242 Novel human DNA-binding protein #8
(R,K) {20,20}
(K) {20}
18: LPSGL KKKKKKKKKKKKKKKKKKK KKK
(K) {20}
19: PGSLL KKKKKKKKKKKKKKKKKKK KK
(K) {20}
20: GSLLK KKKKKKKKKKKKKKKKKKK K
(K) {20}
21: SLKKK KKKKKKKKKKKKKKKKKKK

1
AAU18244 ck: 1109 len: 98 1 Aau18244 Novel human DNA-binding protein #9
(R,K) {20,20}
(K) {20}
53: QTKNT KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
54: TKNTK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: KNTKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
56: NTYKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1
AAU18246 ck: 8102 len: 111 1 Aau18246 Novel human DNA-binding protein #9
(R,K) {20,20}
(K) {20}
78: EFHIL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
79: PHILK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
80: HILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
81: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K) {20}
91: KKKKK KKKKKKKKKKKKKKKKKKK K
(K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKK

1
AAU18247 ck: 8102 len: 111 1 Aau18247 Novel human DNA-binding protein #9
(R,K) {20,20}
(K) {20}
78: EFHIL KKKKKKKKKKKKKKKKKKK KKKKK

79: PHILK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
80: HILKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
81: ILKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
91: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18248 ck: 8319 len: 53 ! Aau18248 Novel human DNA-binding protein #9
(R,K) {20,20}
13: RYFPP KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
14: YFPPK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
15: FPPPK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
16: KPPKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
17: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18250 ck: 7918 len: 80 ! Aau18250 Novel human DNA-binding protein #9
(R,K) {20,20}
50: NVLTV KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
51: VLTVK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
52: LTVKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
53: TVKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
54: VKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18252 ck: 4882 len: 41 ! Aau18252 Novel human DNA-binding protein #9
(R,K) {20,20}
8: FYCFP KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
9: YCFPP KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
10: CFPKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
11: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
12: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

14: KKKK (K) {20} KKKK
15: KKKK (K) {20} KKKK
16: KKKK (K) {20} KKKK
17: KKKK (K) {20} KKKK
18: KKKK (K) {20} KKKK
19: KKKK (K) {20} KKKK
20: KKKK (K) {20} KKKK

AAU18253 ck: 5469 len: 63 ! Aau18253 Novel human DNA-binding protein #1
(R,K) {20,20}
(K) {20}

30: ITCL KKKK (K) {20} KKKK
31: ICLK KKKK (K) {20} KKKK
32: CLLK KKKK (K) {20} KKKK
33: LCLK KKKK (K) {20} KKKK
34: LKKK KKKK (K) {20} KKKK
35: KKKK (K) {20} KKKK
36: KKKK (K) {20} KKKK
37: KKKK (K) {20} KKKK
38: KKKK (K) {20} KKKK
39: KKKK (K) {20} KKKK
40: KKKK (K) {20} KKKK
41: KKKK (K) {20} KKKK
42: KKKK (K) {20} KKKK

AAU18254 ck: 5075 len: 52 ! Aau18254 Novel human DNA-binding protein #1
(R,K) {20,20}
(K) {20}
30: FIVV KKKK (K) {20} KKKK

31: IVVX KKKK (K) {20} KKKK
32: VVXX KKKK (K) {20} KKKK
33: VXXK KKKK (K) {20} KKKK

AAU18255 ck: 5741 len: 47 ! Aau18255 Novel human DNA-binding protein #1
(R,K) {20,20}
(K) {20}

20: ILTF KKKK (K) {20} KKKK
21: LTFK KKKK (K) {20} KKKK
22: TTFK KKKK (K) {20} KKKK
23: TFKK KKKK (K) {20} KKKK
24: FKKK KKKK (K) {20} KKKK
25: KKKK (K) {20} KKKK
26: KKKK (K) {20} KKKK
27: KKKK (K) {20} KKKK

AAU18256 ck: 2868 len: 84 ! Aau18256 Novel human DNA-binding protein #1
(R,K) {20,20}
(K) {20}

53: KCTE KKKK (K) {20} KKKK
54: CTYE KKKK (K) {20} KKKK
55: TYEK KKKK (K) {20} KKKK
56: YEKK KKKK (K) {20} KKKK
57: EKKK (K) {20} KKKK
58: KKKK (K) {20} KKKK
59: KKKK (K) {20} KKKK
60: KKKK (K) {20} KKKK
61: KKKK (K) {20} KKKK
62: KKKK (K) {20} KKKK

AAU18257 ck: 4686 len: 73 ! Aau18257 Novel human DNA-binding protein #1

1 (R,K){20,20}

(K){20}

41: YLKE KKKKKKKKKKKKKKKKKKK

42: LKKE KKKKKKKKKKKKKKKKKKK

(K){20}

43: KKEK KKKKKKKKKKKKKKKKKKK

(K){20}

44: KEKK KKKKKKKKKKKKKKKKKKK

(K){20}

45: EKKE KKKKKKKKKKKKKKKKKKK

(K){20}

46: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

47: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

48: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

49: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

50: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

51: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

52: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

53: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

54: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18258 ck: 6676 len: 74 1 Aau18258 Novel human DNA-binding protein #1

(R,K){20,20}

(K){20}

47: LRTQ KKKKKKKKKKKKKKKKKKK

(K){20}

48: RTQK KKKKKKKKKKKKKKKKKKK

(K){20}

49: TQK KKKKKKKKKKKKKKKKKKK

(K){20}

50: FQK KKKKKKKKKKKKKKKKKKK

(K){20}

51: QKK KKKKKKKKKKKKKKKKKKK

(K){20}

52: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18259 ck: 2283 len: 54 1 Aau18259 Novel human DNA-binding protein #1

(R,K){20,20}

(K){20}

32: IVCF KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18260 ck: 7503 len: 74 1 Aau18260 Novel human DNA-binding protein #1

(R,K){20,20}

(K){20}

33: VCFK KKKKKKKKKKKKKKKKKKK

(K){20}

34: PCFK KKKKKKKKKKKKKKKKKKK

(K){20}

33: VCFK KKKKKKKKKKKKKKKKKKK

(K){20}

34: PCFK KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18260 ck: 7503 len: 74 1 Aau18260 Novel human DNA-binding protein #1

(R,K){20,20}

(K){20}

45: SHLD KKKKKKKKKKKKKKKKKKK

(K){20}

46: HLTD KKKKKKKKKKKKKKKKKKK

(K){20}

47: LTDK KKKKKKKKKKKKKKKKKKK

(K){20}

48: TDKK KKKKKKKKKKKKKKKKKKK

(K){20}

49: DKKK KKKKKKKKKKKKKKKKKKK

(K){20}

50: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

51: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

52: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

53: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

54: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

55: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18262 ck: 5199 len: 84 1 Aau18262 Novel human DNA-binding protein #1

(R,K){20,20}

(K){20}

63: ANNA KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18263 ck: 7578 len: 31 1 Aau18263 Novel human DNA-binding protein #1

(R,K){20,20}

(K){20}

6: LTEL KKKKKKKKKKKKKKKKKKK

(K){20}

7: TELK KKKKKKKKKKKKKKKKKKK

(K){20}

8: ELEK KKKKKKKKKKKKKKKKKKK

(K){20}

9: LEKK KKKKKKKKKKKKKKKKKKK

(K){20}

10: EKKK KKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKK KKKKKKKKKKKKKKKKKKK

(K){20}

AAU18264 ck: 3915 len: 57 1 Aau18264 Novel human DNA-binding protein #1


```
1
  (R,K){20,20}
31: KQLLL KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
32: QLLLK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
33: LLLKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20} XGGF

AAU18265 ck: 3679 len: 37 ! Aau18265 Novel human DNA-binding protein #1
  (R,K){20,20}
15: ISPLT KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
16: SPLTK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
17: PLTKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20} X

AAU18266 ck: 657 len: 196 ! Aau18266 Novel human DNA-binding protein #1
  (R,K){20,20}
169: FVYFE KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
170: VYFEK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
171: XFPEK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
172: FEKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
173: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20} X

AAU18267 ck: 4672 len: 57 ! Aau18267 Novel human DNA-binding protein #1
  (R,K){20,20}
28: DKTFH KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
29: KTFHK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
30: TFHKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
31: FHKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
32: HKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20} XPGG
```

```
1
AAU18268 ck: 9656 len: 66 ! Aau18268 Novel human DNA-binding protein #1
  (R,K){20,20}
38: WVISV KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
39: VISVK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
40: ISVKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
41: SVKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
42: VKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (R,K){20} EXKK

AAU18270 ck: 4665 len: 34 ! Aau18270 Novel human DNA-binding protein #1
  (R,K){20,20}
10: PELLL KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
11: ELLLK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
12: LLLKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
13: LLKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
14: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAU18271 ck: 7810 len: 64 ! Aau18271 Novel human DNA-binding protein #1
  (R,K){20,20}
37: LKYPW KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
38: KYFWK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
39: YFWKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
40: FWKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
41: WKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20} XGXP

AAU18272 ck: 269 len: 76 ! Aau18272 Novel human DNA-binding protein #1
  (R,K){20,20}
37: TPSRA KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
38: PSRAK KKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}
```

39: SRKKK (K) {20}
 40: RAKKK (K) {20}
 41: AKKKK (K) {20}
 42: KKKKK (K) {20}
 43: KKKKK (K) {20}
 44: KKKKK (K) {20}

AAU18273 ck: 8370 len: 45 i Aau18273 Novel human DNA-binding protein #1

17: APTQK (R,K) {20,20}
 18: PRTQK (K) {20}
 19: KTQKK (K) {20}
 20: TQKKK (K) {20}
 21: QKKKK (K) {20}
 22: KKKKK (K) {20}
 23: KKKKK (K) {20}
 24: KKKKK (K) {20}
 25: KKKKK (K) {20}

AAU18274 ck: 1663 len: 87 i Aau18274 Novel human DNA-binding protein #1

36: KWSXX (R,K) {20,20}
 37: WSSXX (K) {20}
 38: SSXXX (K) {20}
 39: SXKKK (K) {20}
 40: XXXKK (K) {20}
 41: KKKKK (K) {20}

42: KKKKK
 43: KKKKK (K) {20}
 44: KKKKK (K) {20}
 45: KKKKK (K) {20}
 46: KKKKK (K) {20}
 47: KKKKK (K) {20}
 48: KKKKK (K) {20}
 49: KKKKK (K) {20}
 50: KKKKK (K) {20}
 51: KKKKK (K) {20}
 52: KKKKK (K) {20}
 53: KKKKK (K) {20}
 54: KKKKK (K) {20}
 55: KKKKK (K) {20}
 56: KKKKK (K) {20}
 57: KKKKK (K) {20}
 58: KKKKK (K) {20}
 59: KKKKK (K) {20}
 60: KKKKK (K) {20}
 61: KKKKK (K) {20}
 62: KKKKK (K) {20}
 63: KKKKK (R,K) {20}

AAU18275 ck: 5607 len: 63 i Aau18275 Novel human DNA-binding protein #1

26: MVELK (R,K) {20,20}
 (K) {20}

27:	VELK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
28:	ELBK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
29:	LEKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
30:	EXKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
31:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
32:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
33:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
34:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
35:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
36:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
37:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
38:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
39:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	
40:	KKKK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXX
		(X) { 20 }	

```

AAU18276 ck: 5997 len: 58      1 Aau18276 Novel human DNA-binding protein #1
      (R,K){20,20}
      (K){20}
28:  RPTRP  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
29:  PTRPK  KKKKKKKKKKKKKKKKKKKKK KKKKK

```

```

1
AAU18277   ck: 5764    len: 74      | Aau18277 Novel human DNA-binding protein #1
          (R,K){20,20}
40: EFLSA  KKKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}
41: FLSAK  KKKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}
42: LSAKK  KKKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}
43: SAKKK  KKKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}
44: AKKKK  KKKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}
45: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}

```

```

46: KKKK {X} {20} KKKK
47: KKKK {X} {20} KKKK
48: KKKK {X} {20} KKKK
49: KKKK {X} {20} KKKK
50: KKKK {X} {20} KKKK
51: KKKK {X} {20} KKKK
52: KKKK {X} {20} KKKK
53: KKKK {X} {20} KK
54: KKKK {X} {20} X

```

```

1
AAU18278 ck: 7734 len: 97 1 Aau18278 Novel human DNA-binding protein #1
(R,X){20,20}
57: RGCSY KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
58: GCSYK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
59: CSYK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
60: SYKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
61: YKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
62: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
63: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
64: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
65: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
66: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
67: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
68: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
69: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}
70: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(X){20}

```

71: KKKKX {20} KKKKX
 (X) {20}
 72: KKKKX {20} KKKKG
 (X) {20}
 73: KKKKX {20} KKKGX
 (X) {20}
 74: KKKKX {20} KKKGX
 (X) {20}
 75: KKKKX {20} KKKGX

AAU18279 CK: 3164 len: 181 | Aau18279 Novel human DNA-binding protein #1

(R,K) {20,20}

154: TRKPE KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

155: RKPER KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

156: KPERR KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

157: PEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

158: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

159: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

160: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
 (K) {20}

161: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK X
 (K) {20}

```

AAU18280 ck: 7117 len: 55 | Aau18280 Novel human DNA-binding protein #1
                                     (R,K){20,20}
22: DDKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
23: DKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
26: NKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

```

```

1      AAU18281  ck: 9316  len: 67      | Aaul8281 Novel human DNA-binding protein #1
      (R,K){20,20}
      (R,K){20}
      36: FTIQ  RRRRRRRRRRRRRRRRRR  KKKKK
      (K){20}

```

```

37: TLQTR KKKKKKKKKKKKKKKKKKKKK KKKKK
      (X) { 20 }

38: LQTRK KKKKKKKKKKKKKKKKKKKKK KKKKG
      (X) { 20 }

39: QTRKK KKKKKKKKKKKKKKKKKKKKK KKKGG
      (X) { 20 }

40: TRKKK KKKKKKKKKKKKKKKKKKKKK KKGGR
      (X) { 20 }

41: RKKKK KKKKKKKKKKKKKKKKKKKKK XGGNR
      (X) { 20 }

```

1

AAM53862 ck: 5383 len: 86 ! Aam53862 Human brain expressed single exon

15: RRRRG RRRRRKKKKKKKKKKKKKK KRRRR
(R,X){20}
(R,X){20}

16: RRRGR RRRRKKKKKKKKKKKKKK KRRRR
(R,X){20}

17: RRRGR RRRKKKKKKKKKKKKKK KRRRR
(R,X){20}

18: RGRRR RRRKKKKKKKKKKKKKK KRRRR
(R,X){20}

19: GRRRR RKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

20: RRRRR KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

21: RRRRK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

22: RRRKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

23: RRRKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

24: RRRKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

25: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

26: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

27: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

28: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

29: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

30: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

31: RKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

32: KKKKK KKKKKKKKKKKKKKKKK RRRRR
(R,X){20}

33: KKKKK KKKKKRRRRRRRRRRRRRR RRRRR
 (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRR RRRRR
 (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRR RRRRR
 (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRR RRRRR
 (R){20}
37: RKKKK RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
38: KKKKK RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
39: KKKRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
40: KKKRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
41: KRRRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
42: RRRRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
43: RRRRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
44: RRRRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}
45: RRRRR RRRRRRRRRRRRRRRRR RRRRR
 (R){20}

AAM54700 ck: 1334 len: 86 | Aam54700 Human brain expressed single exon

57: EEEEG RRKKKKRRKKKKKKKKKK
 (R,K){20,
 (R,K){20}
58: EEGRR RRKKKKRRKKKKKKKKKK
 (R,K){20}
59: EGGRR RRKKKKRRKKKKKKKKKK
 (R,K){20}
60: EGRRR KKKRRKKKKKKKKKKKKKK
 (R,K){20}
61: GRRRK KKKRRKKKKKKKKKKKKKK
 (R,K){20}
62: RRRKK KKKRRKKKKKKKKKKKKKK
 (R,K){20}
63: RRRKK KRRKKKKKKKKKKKKKKKK
 (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKK
 (R,K){20}
65: KKKRR KRRKKKKKKKKKKKKKKKK
 (R,K){20}
66: KKKRR RKKKKKKKKKKKKKKKKKK
 (K){20}

67: KKKKK KKKKKKKKKKKKKKKKKKKKK

1 AAM54795 ck: 9082 len: 167 ! Aam54795 Human brain expressed single exon
(R,K){20,20}
33: EBGKG RRRRRRRKKRRRRKKKK RKGKG
(R,K){20}
34: EGGGR RRRRRRRKKRRRRKKKK RGGGR
(R,K){20}
35: GGGRR RRRRRRRKKRRRRKKKK RGGRR
(R,K){20}

1 AAM55320 ck: 2276 len: 89 ! Aam55320 Human brain expressed single exon
(R,K){20,20}
23: EEEEE KKKKKKKKKKKKKKKKKKK EEEK
(R,K){20}
46: KKEE KKKKKKKKKKKKKKKKKKK KKEE
(K){20}
47: KEEK KKKKKKKKKKKKKKKKKKK KKEE
(K){20}
48: EEEK KKKKKKKKKKKKKKKKKKK KEEE
(R,K){20}
49: EEEK KKKKKKKKKKKKKKKKKKK EEEE
(R,K){20}

1 AAM55623 ck: 1939 len: 130 ! Aam55623 Human brain expressed single exon
(R,K){20,20}
42: EGGKE RRRRRRRKKRRRRKKKK RRRR
(R,K){20}
43: GKKR RRRRRRRKKRRRRKKKK RRRR
(R,K){20}
44: RKEER RRRRRRRKKRRRRKKKK RRRR
(R,K){20}
45: KEERR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}
46: EEEEE RRRRRRRKKRRRRKKKK RRRK
(R,K){20}
47: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}
48: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}
49: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}

50: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}

51: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}

52: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}

53: RRRR RRRRRRRKKRRRRKKKK RRRK
(R,K){20}

1

AAM55931 ck: 3607 len: 88 i Aam55931 Human brain expressed single exon

(R,K){20}

54: RRRRR RRRRRRRRRRRRRRRRRRR EEEEE

(R,K){20,20}

39: ERRRE KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

40: RKREK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KRREK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: REKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

1

```
61: KKKK (K) {20} KKKK
62: KKKK (K) {20} KKKK
63: KKKK (K) {20} KKKK
64: KKKK (K) {20} KKKK
65: KKKK (K) {20} KKKK
66: KKKK (K) {20} KKKK
67: KKKK (K) {20} KKKK
68: KKKK (K) {20} KKKK
69: KKKK (K) {20} KKKK

AAMS6490 ck: 3937 len: 85 i Aams6490 Human brain expressed single exon i
1: KKKK (R, K) {20, 20} KKKK
2: KKKK (K) {20} KKKK
3: KKKK (K) {20} KKKK
4: KKKK (K) {20} KKKK
5: KKKK (K) {20} KKKK
6: KKKK (K) {20} KKKK
7: KKKK (K) {20} KKKK
8: KKKK (K) {20} KKKK
9: KKKK (K) {20} KKKK
10: KKKK (K) {20} KKKK
11: KKKK (K) {20} KKKK

AAMS8224 ck: 1560 len: 88 i Aams8224 Human brain expressed single exon i
1: KKKK (R, K) {20, 20} KKKK
43: KKKK (R, K) {20} KKKK
```

```

44: ERGR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
45: RGR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
46: RGR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
47: GRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
48: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
49: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
50: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
51: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
52: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
53: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
54: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
55: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
56: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
57: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
58: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
59: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
60: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
61: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
62: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
AAM58395 ck: 2324 len: 36 ! Aam58395 Human brain expressed single exon
    (R,K){20,20}
    (R,K){20}
16: KERKT KKKKKRRRRRRKKRRKKRR R
    (R,K){20}
17: ERKTK KKKKKRRRRRRKKRRKKRR
    (R,K){20}
AAM61063 ck: 8343 len: 66 ! Aam61063 Human brain expressed single exon
    (R,K){20,20}
    (R,K){20}
6: ETERE KKKKKKKKKKKKKKKKKKK KKKK

```

```

7: TEREX KKKKKKKKKKKKKKKKKKK KKKK
    (R,K){20}
8: BREKK KKKKKKKKKKKKKKKKKKK KKKK
    (R,K){20}
9: REKKK KKKKKKKKKKKKKKKKKKK NKKK
    (R,K){20}
AAM63533 ck: 2394 len: 57 ! Aam63533 Human brain expressed single exon
    (R,K){20,20}
    (R){20}
20: EEEEG RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
21: EEEGR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
22: EEEGR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
23: EGRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
24: GRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
25: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR
    (R){20}
AAM64090 ck: 4228 len: 24 ! Aam64090 Human brain expressed single exon
    (R,K){20,20}
    (R,K){20}
1: RRRRRKKKKRRRRRRRRRR RRKT
    (R,K){20}
2: R RRRRRKKKKRRRRRRRRRR RRKT
    (R,K){20}
3: RR RRRRRKKKKRRRRRRRRRR RR
    (R,K){20}
4: RRR RRRRRKKKKRRRRRRRRRR T
    (R,K){20}
AAM64863 ck: 3301 len: 52 ! Aam64863 Human brain expressed single exon
    (R,K){20,20}
    (R,K){20}
12: KKKK KKKKKKKKKRRRRKKKK KKKK
    (R,K){20}
13: KKKK KKKKKKKKKRRRRKKKK KKKK
    (R,K){20}
14: KKKK KKKKKKKKKRRRRKKKK KKKK
    (R,K){20}
15: KKKK KKKKKKKKKRRRRKKKK KKKK
    (R,K){20}
16: NKKK KKKKKRRRRRRKKKKKK KKKK
    (R,K){20}
17: KKKK KKKKKRRRRRRKKKKKK KKKK
    (R,K){20}
18: KKKK KKKKKRRRRRRKKKKKK KKKK
    (R,K){20}

```

19: KKKKK KGRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
20: KKKKK KRRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
21: KKKKK RKRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
22: KKKKK KRRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
23: KKKKK RKRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
24: KKKKK KRRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
25: KRRKK RKRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
26: RRRKK KRRKKRKKKKKKKKKKKK KKKKK
(R,K){20}
27: KRRKK RKKKKRKKKKKKKKKKKK KKKKA
(R,K){20}
28: RRRKK KKKKKRKKKKKKKKKKKK KKKAF
(K){20}
29: KRRKK KKKKKRKKKKKKKKKKKK KKAFF
(K){20}
30: RRRKK KKKKKRKKKKKKKKKKKK KAF
(K){20}
31: KRRKK KKKKKRKKKKKKKKKKKK AF
(K){20}

1
AAM65355 ck: 4895 len: 51 | Aam65355 Human brain expressed single exon
(R,K){20,20}
18: LFKPM RKRKKRKKRRKKKKKKRR KCLTT
(R,K){20}
19: FKPMR KRRKKRKKRRKKKKRRR KLT TT
(R,K){20}
20: KPMRK RRRKKRKKRRKKKKRRR LTT TT
(R,K){20}

1
AAM66249 ck: 5383 len: 86 | Aam66249 Human bone marrow expressed probe
(R,K){20,20}
15: RRRRG RRRRKKKKKKKKKKKKKK KKR RR
(R,K){20}
16: RRRGR RRRRKKKKKKKKKKKKKK KRR RR
(R,K){20}
17: RRGRR RRRRKKKKKKKKKKKKKK RRR RR
(R,K){20}
18: RGRRR RRRKKKKKKKKKKKKKK RRR RR
(R,K){20}
19: GRRRR RKKKKRKKKKKKKKKKRR RRR RR
(R,K){20}
20: RRRRR KKKKKRKKKKKKKKRRR RRR RR
(R,K){20}

21: RRRRK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
22: RRRKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
23: RRRKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
24: RKKKK KRRKKRKKKKKKKKRRR RRRR
(R,K){20}
25: KKKKK KRRKKRKKKKKKKKRRR RRRR
(R,K){20}
26: KKKKK RKKKKRKKKKKKKKRRR RRRR
(R,K){20}
27: KKKKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
28: KKKKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
29: KKKKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
30: KKKKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
31: RKKKK KRRKKRKKKKKKKKRRR RRRR
(R,K){20}
32: KKKKK RKKKKRKKKKKKKKRRR RRRR
(R,K){20}
33: KKKKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
34: KKKKK KKKKKRKKKKKKKKRRR RRRR
(R,K){20}
35: KKKKK KRRKKRKKKKKKKKRRR RRRR
(R,K){20}
36: KKKKK KRRKKRKKKKKKKKRRR RRRR
(R,K){20}
37: RKKKK RRRRRRKKKKKKKKRRR RRRR
(R){20}
38: KKKKK RRRRRRKKKKKKKKRRR RRRR
(R){20}
39: KKKKK RRRRRRKKKKKKKKRRR RRRR
(R){20}
40: KKKKK RRRRRRKKKKKKKKRRR RRRR
(R){20}
41: KRRRR RRRRRRKKKKKKKKRRR RRRR
(R){20}
42: RRRRR RRRRRRKKKKKKKKRRR RRRR
(R){20}
43: RRRRR RRRRRRKKKKKKKKRRR RRRR
(R){20}
44: RRRRR RRRRRRKKKKKKKKRRR RRRR
(R){20}


```

45: RRRR RRRRRRRRRRRRRRRR NKQT
AAM67100 ck: 1334 len: 86 ! Aam67100 Human bone marrow expressed probe
1
57: EEEEG RRRRRRRRRRRRRRRR (R,K){20,20}
(R,K){20}
58: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
59: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
60: EGRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
61: GRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
62: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
63: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
64: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
65: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
66: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
67: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
AAM67180 ck: 9082 len: 167 ! Aam67180 Human bone marrow expressed probe
1
33: EEEGR RRRRRRRRRRRRRRRR (R,K){20,20}
(R,K){20}
34: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
35: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
AAM6717 ck: 2276 len: 89 ! Aam6717 Human bone marrow expressed probe
1
23: EEEEG RRRRRRRRRRRRRRRR (R,K){20,20}
(R,K){20}
46: EEEEG RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
47: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
48: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
49: EEEGR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
AAM68007 ck: 1939 len: 130 ! Aam68007 Human bone marrow expressed probe

```

```

1
42: EGRKE RRRRRRRRRRRRRRRR (R,K){20,20}
(R,K){20}
43: GRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
44: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
45: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
46: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
47: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
48: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
49: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
50: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
51: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
52: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
53: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
54: RRRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
AAM68298 ck: 3607 len: 88 ! Aam68298 Human bone marrow expressed probe
1
39: ERRR RRRRRRRRRRRRRRRR (R,K){20,20}
(R,K){20}
40: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
41: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
42: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
43: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
44: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
45: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
46: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}
47: ERRR RRRRRRRRRRRRRRRR (R,K){20}
(R,K){20}

```

```

48: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
49: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
50: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
51: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
52: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
53: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
54: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
55: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
56: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
57: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
58: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
59: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
60: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
61: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
62: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
63: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
64: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
65: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
66: KKKK KKKKKKKKKKKKKKKKKKK KKK
    (K) {20}
67: KKKK KKKKKKKKKKKKKKKKKKK KK
    (K) {20}
68: KKKK KKKKKKKKKKKKKKKKKKK K
    (K) {20}
69: KKKK KKKKKKKKKKKKKKKKKKK

AAM6869 ck: 3937 len: 85 1 Aam6869 Human bone marrow expressed probe
(R, K) {20,20}
(K) {20}
1: KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}

```

```

2: K KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
3: KK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
4: KK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
5: KKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
6: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
8: KKKK KKKKKKKKKKKKKKKKKKK KKEE
    (K) {20}
9: KKKK KKKKKKKKKKKKKKKKKKK KKEE
    (K) {20}
10: KKKK KKKKKKKKKKKKKKKKKKK KEEE
    (K) {20}
11: KKKK KKKKKKKKKKKKKKKKKKK EEEE

AAM69569 ck: 2686 len: 71 1 Aam69569 Human bone marrow expressed probe
(R, K) {20,20}
(K) {20}
20: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
21: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (K) {20}
22: KKOK KKKKKKKKKKKKKKKKKKK KKKR
    (K) {20}
23: KOKK KKKKKKKKKKKKKKKKKKK KRRK
    (K) {20}
24: OKKK KKKKKKKKKKKKKKKKKKK RRRK
    (R, K) {20}
25: KKKK KKKKKKKKKKKKKKKKKKK KRRK
    (R, K) {20}
26: KKKK KKKKKKKKKKKKKKKKKKK RKKK
    (R, K) {20}
27: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (R, K) {20}
28: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (R, K) {20}
29: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (R, K) {20}
30: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (R, K) {20}
31: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (R, K) {20}
32: KKKK KKKKKKKKKKKKKKKKKKK KKKK
    (R, K) {20}

```

```

33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}

AAM70678 ck: 1560 len: 88 ! Aam70678 Human bone marrow expressed probe
      (R,K) {20,20}
43: RRRRG RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
44: RRRGR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
45: RRRGR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
46: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
47: GRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
48: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}

```

```

49: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
50: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
51: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
52: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
53: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
54: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
55: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
56: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
57: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
58: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
59: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
60: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
61: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
62: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}

AAM70881 ck: 2324 len: 36 ! Aam70881 Human bone marrow expressed probe
      (R,K) {20,20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}

AAM7367 ck: 8343 len: 66 ! Aam7367 Human bone marrow expressed probe
      (R,K) {20,20}
6: ETERE KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
7: TEREK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
8: EREKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}
9: REKKK KKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20}

AAM76347 ck: 2394 len: 57 ! Aam76347 Human bone marrow expressed probe
      (R,K) {20,20}
20: EEEEG RRRRRRRRRRRRRRRRRRR RRRRR
      (R) {20}

```

1

21: EEEGR RRRRRRRRRRRRRRRRRR RRRRG (R) {20}

22: EGGRR RRRRRRRRRRRRRRRRRR RRRGR (R) {20}

23: EGGRR RRRRRRRRRRRRRRRRRR RRRGR (R) {20}

24: GRRRR RRRRRRRRRRRRRRRRRR RRRRR (R) {20}

25: RRRRR RRRRRRRRRRRRRRRRRR GRRRR (R) {20}

1

AAM78048 ck: 4228 len: 24 ! Aam76911 Human bone marrow expressed probe

1: (R, K) {20, 20}

(R, K) {20}

RRRRRRRRRRRRRRRRRR RRRRT

2: (R, K) {20}

R RRRRRRRRRRRRRRRR RRT

3: (R, K) {20}

RRRRRRRRRRRRRRRRR RT

4: RRR RRRRRRRRRRRRRRRR T

1

AAM78048 ck: 4895 len: 51 ! Aam78048 Human bone marrow expressed probe

(R, K) {20, 20}

(R, K) {20}

18: LFKPM RRRRRRRRRRRRRRRR KKLTT

19: FKPMR RRRRRRRRRRRRRR KLTTT (R, K) {20}

20: KPMRK RRRRRRRRRRRRRR LTTT (R, K) {20}

1

AAM82533 ck: 1736 len: 40 ! Aam82533 Human immune/haematopoietic antigen

(R, K) {20, 20}

(K) {20}

18: LPSL KKKKKKKKKKKKKKKK KKK (K) {20}

19: PPSLK KKKKKKKKKKKKKK KK (K) {20}

20: GSLK KKKKKKKKKKKKKK K (K) {20}

21: SLKK KKKKKKKKKKKKKK (K) {20}

1

AAM85748 ck: 7503 len: 74 ! Aam85748 Human immune/haematopoietic antigen

(R, K) {20, 20}

(K) {20}

45: SHLTD KKKKKKKKKKKKKKKK KKKK (K) {20}

46: HLTDK KKKKKKKKKKKKKKKK KKKK (K) {20}

47: LTDDK KKKKKKKKKKKKKKKK KKKK (K) {20}

```

48: TDKKK KKKKKKKKKKKKKKKKKKK (K) {20}
49: DKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}

AAM90546 ck: 6676 len: 74 ! Aam90546 Human immune/haematopoietic antigen
(R, K) {20,20}
47: LRTFQ KKKKKKKKKKKKKKKKKKK (K) {20}
48: RTFQK KKKKKKKKKKKKKKKKKKK (K) {20}
49: TFOKK KKKKKKKKKKKKKKKKKKK (K) {20}
50: FOKKK KKKKKKKKKKKKKKKKKKK (K) {20}
51: OKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}

AAM90618 ck: 5691 len: 108 ! Aam90618 Human immune/haematopoietic antigen
(R, K) {20,20}
78: VRPCL KKKKKKKKKKKKKKKKKKK (K) {20}
79: RPCLK KKKKKKKKKKKKKKKKKKK (K) {20}
80: PCLKK KKKKKKKKKKKKKKKKKKK (K) {20}
81: CLKKK KKKKKKKKKKKKKKKKKKK (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20}

```

85: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKK
 (K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKK KK
 (K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKK K
 (K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKK

1
AAM91030 ck: 8102 len: 111 ! Aam91030 Human immune/haematopoietic antigen
(R,K) {20,20}

78: EPHIL KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
79: PHILK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
80: HILKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
81: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKK KKK
 (K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKK KK
 (K) {20}
91: KKKKK KKKKKKKKKKKKKKKKKKK K
 (K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKK

1
AAM91162 ck: 1109 len: 98 ! Aam91162 Human immune/haematopoietic antigen
(R,K) {20,20}

53: QTKNT KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
54: TQNTK KKKKKKKKKKKKKKKKKKK KKKKK

1

55: KNTTK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
56: NTKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
57: TKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAM91891 ck: 8102 len: 111 ! Aam91891 Human immune/haematopoietic antigen
(R,K) {20,20}

78: EPHIL KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
79: PHILK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
80: HILKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
81: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

87: KKKK (K) {20}
88: KKKK (K) {20}
89: KKKK (K) {20}
90: KKKK (K) {20}
91: KKKK (K) {20}
92: KKKK (K) {20}

AA091933 ck: 8285 len: 55 ! Aa091933 Human digestive system antigen SEQ

1

33: PTPRP (R,K) {20,20}
34: PTPRP (K) {20}

33: PTPRP KKKK (K) {20}
34: PTPRP KKKK (K) {20}

AA092433 ck: 1663 len: 87 ! Aa092433 Human digestive system antigen SEQ

1

36: KWSX (R,K) {20,20}
37: WSSX (K) {20}
38: SSXX (K) {20}
39: SXKK (K) {20}
40: XKKK (K) {20}
41: KKKK (K) {20}
42: KKKK (K) {20}
43: KKKK (K) {20}
44: KKKK (K) {20}
45: KKKK (K) {20}
46: KKKK (K) {20}
47: KKKK (K) {20}
48: KKKK (K) {20}
49: KKKK (K) {20}

1

50: KKKK (K) {20}
51: KKKK (K) {20}
52: KKKK (K) {20}
53: KKKK (K) {20}
54: KKKK (K) {20}
55: KKKK (K) {20}
56: KKKK (K) {20}
57: KKKK (K) {20}
58: KKKK (K) {20}
59: KKKK (K) {20}
60: KKKK (K) {20}
61: KKKK (K) {20}
62: KKKK (K) {20}
63: KKKK (R,K) {20}

AA000092 ck: 9065 len: 113 ! Aa000092 Human polypeptide SEQ ID NO 13984.

1

91: CLGCL (R,K) {20,20}
92: LGCLK (K) {20}

91: CLGCL KKKK (K) {20}
92: LGCLK KKKK (K) {20}

AA000222 ck: 2916 len: 132 ! Aa000222 Human polypeptide SEQ ID NO 14114.

1

28: XPLPP (R,K) {20,20}

28: XPLPP KKKK (K) {20}

AA000232 ck: 1000 len: 102 ! Aa000232 Human polypeptide SEQ ID NO 14124.

1

29: RDCFP (R,K) {20,20}
30: DCFPK (K) {20}
31: CFFKK (K) {20}

29: RDCFP KKKK (K) {20}
30: DCFPK KKKK (K) {20}
31: CFFKK KKKK (K) {20}

32:	FPKKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
33:	FKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
34:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
35:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
36:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
37:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
38:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
39:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
40:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
41:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
42:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
43:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
44:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
45:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
46:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
47:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
48:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
49:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
50:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
51:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
52:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKKKK
		(X) {20}	
53:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KKGGP
		(X) {20}	
54:	KKKKK	XXXXXXXXXXXXXXXXXXXX	KGPP
		(X) {20}	
55:	KKKKK	XXXXXXXXXXXXXXXXXXXX	GGPP
		(X) {20}	

```

1      AA000291 ck: 8100 len: 124   ! Aa000291 Human polypeptide SEQ ID NO 14183.
          (R,K){20,20}
          (K){20}
22: QHFCM KKKKKKKKKKKKKKKKKKKKK FFKKG

1      AA000439 ck: 6396 len: 122   ! Aa000439 Human polypeptide SEQ ID NO 14331.
          (R,X){20,20}
          (X){20}
23: CLMLV KKKKKKKKKKKKKKKKKKKKK KKKKK
24: LMLVK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
25: WLVRK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
26: LVRRK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
27: VRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (R,X){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (R,X){20}

```


51: KKKK (K) {20} KKKK
52: KKKK (K) {20} KKKK
53: KKKK (K) {20} KKKK
54: KKKK (K) {20} KKKK Q
AA002186 ck: 4844 len: 57 ! Aa002186 Human polypeptide SEQ ID NO 16078.
(R, K) {20, 20}
11: HCCL KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
12: CCLL KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
AA002310 ck: 704 len: 137 ! Aa002310 Human polypeptide SEQ ID NO 16202.
(R, K) {20, 20}
(R, K) {20}
22: HSLN KKKKKKKKKKKKKKKKKKKKK KKKK
AA002477 ck: 2018 len: 112 ! Aa002477 Human polypeptide SEQ ID NO 16369.
(R, K) {20, 20}
(K) {20}
83: LASAV KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
84: ASAV KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
85: SAVK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
86: AVKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
87: VKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
88: KKKK (K) {20} KKKK
89: KKKK (K) {20} KKKK
90: KKKK (K) {20} KKKK
91: KKKK (K) {20} KKKK
92: KKKK (K) {20} KKKK
93: KKKK (K) {20} KKKK
AA002733 ck: 2860 len: 132 ! Aa002733 Human polypeptide SEQ ID NO 16625.
(R, K) {20, 20}
(K) {20}
86: FFFSL KKKKKKKKKKKKKKKKKKKKK KKKK

1
87: FFLK (K) {20} KKKK
88: FSLK (K) {20} KKKK
89: SLKK (K) {20} KKKK
90: LKKK (K) {20} KKKK
91: KKKK (K) {20} KKKK
92: KKKK (K) {20} KKKK
93: KKKK (K) {20} KKKK
94: KKKK (K) {20} KKKK
95: KKKK (K) {20} KKKK
96: KKKK (K) {20} KKKK
97: KKKK (K) {20} KKKK
98: KKKK (K) {20} KKKK
99: KKKK (K) {20} KKKK
100: KKKK (K) {20} KKKK
101: KKKK (R, K) {20} KKKK
AA002946 ck: 3649 len: 126 ! Aa002946 Human polypeptide SEQ ID NO 16838.
(R, K) {20, 20}
(K) {20}
25: DEATS KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
26: EATSK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
27: ATSK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
28: TSKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
29: SKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
30: KKKK (K) {20} KKKK
31: KKKK (K) {20} KKKK
32: KKKK (K) {20} KKKK

33: KKKK (K) {20}
 34: KKKK (K) {20}
 35: KKKK (K) {20}
 36: KKKK (K) {20}
 37: KKKK (K) {20}
 38: KKKK (K) {20}

AA002961 ck: 4320 len: 83 i Aa002961 Human polypeptide SEQ ID NO 16853.
 22: RMFSS KKKK (R,K) {20,20} KTAIT
 23: MFSSK KKKK (K) {20} TAITK

AA003006 ck: 5212 len: 102 i Aa003006 Human polypeptide SEQ ID NO 16898.
 59: HDFFP KKKK (R,K) {20,20} KKKK
 60: DFFPK KKKK (K) {20} KKKK
 61: FFFPK KKKK (K) {20} KKKK
 62: FPKKK KKKK (K) {20} KKKK
 63: PPKKK KKKK (K) {20} KKKK
 64: KKKKK KKKK (K) {20} KKKK
 65: KKKKK KKKK (K) {20} KKKK
 66: KKKKK KKKK (K) {20} KKKK
 67: KKKKK KKKK (K) {20} ASSSS
 AA003024 ck: 6865 len: 64 i Aa003024 Human polypeptide SEQ ID NO 16916.
 15: SPAKA KKKK (R,K) {20,20} KRGFP
 16: PAKAR KKKK (K) {20} RGGP
 17: AKAPK KKKK (R,K) {20} GGPPL

AA003113 ck: 5587 len: 60 i Aa003113 Human polypeptide SEQ ID NO 17005.
 33: EIMNL KKKK (R,K) {20,20} KKKR
 34: INMLK KKKK (K) {20} KRG
 35: NMLKK KKKK (K) {20} KRG
 36: MLKKK KKKK (K) {20} RGG
 37: LKKKK KKKK (R,K) {20} GGL

AA003123 ck: 6627 len: 108 i Aa003123 Human polypeptide SEQ ID NO 17015.
 33: RTWRX KKKK (R,K) {20,20} KKKK
 34: TWRRX KKKK (K) {20} KKKK
 35: WRXXK KKKK (K) {20} KKKK
 36: RXKKK KKKK (K) {20} KKKK
 37: XXXKK KKKK (K) {20} KKKK
 38: KKKKK KKKK (K) {20} KKKR
 39: KKKKK KKKK (K) {20} KKR
 40: KKKKK KKKK (K) {20} KRG
 41: KKKKK KKKK (K) {20} KRG
 42: KKKKK KKKK (K) {20} RGG
 43: KKKKK KKKK (R,K) {20} GGGK

AA003132 ck: 3903 len: 116 i Aa003132 Human polypeptide SEQ ID NO 17024.
 1: KKKK (R,K) {20,20} KKKR
 2: KKKKK KKKK (K) {20} KKR
 3: KK KKKK (K) {20} KRG
 4: KKK KKKK (K) {20} KRG
 5: KKK KKKK (K) {20} RGGP

(R,K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKR GGPPK

AAO03152 ck: 7891 len: 35 i Aao03152 Human polypeptide SEQ ID NO 17044.
(R,K){20,20}
(R,K){20}
13: LIXYT RKKKKKKKKKKKKKKKKKKR RGG
(R,K){20}
14: IXYTR KKKKKKKKKKKKKKKKKRR GG

AAO03168 ck: 7233 len: 67 i Aao03168 Human polypeptide SEQ ID NO 17060.
(R,K){20,20}
(K){20}
17: ELMP KKKKKKKKKKKKKKKKKKKR KRGA
(K){20}
18: LMMP KKKKKKKKKKKKKKKKKKKR RGGAP
(R,K){20}
19: LMPK KKKKKKKKKKKKKKKKKRR CGAPF

AAO03243 ck: 411 len: 95 i Aao03243 Human polypeptide SEQ ID NO 17135.
(R,K){20,20}
(K){20}
51: LHAV KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
52: LHAV KKKKKKKKKKKKKKKKKKKR KKKR
(K){20}
53: HAVK KKKKKKKKKKKKKKKKKRR EKKRG

AAO03273 ck: 2981 len: 88 i Aao03273 Human polypeptide SEQ ID NO 17165.
(R,K){20,20}
(K){20}
43: IXYL KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
44: XYLL KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
45: YLLK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
46: LLKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
47: LKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
48: KKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
49: KKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
50: KKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
51: KKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
52: KKKK KKKKKKKKKKKKKKKKKKKR KKKK

(K){20}
53: KKKK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
54: KKKK KKKKKKKKKKKKKKKKKKKR KKKPT
(K){20}
55: KKKK KKKKKKKKKKKKKKKKKKKR KKPPT
(K){20}
56: KKKK KKKKKKKKKKKKKKKKKKKR KPTPK
(K){20}
57: KKKK KKKKKKKKKKKKKKKKKKKR PTPK

AAO03277 ck: 2981 len: 89 i Aao03277 Human polypeptide SEQ ID NO 17169.
(R,K){20,20}
(K){20}
1: KKKKKKKKKKKKKKKKKKKR KKKRM
(K){20}
2: K KKKKKKKKKKKKKKKKKKKR KKKMK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKR KKKMN
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKR KKKNK
(R,K){20}
5: KKK KKKKKKKKKKKKKKKKKRR MKNK

AAO03278 ck: 6777 len: 74 i Aao03278 Human polypeptide SEQ ID NO 17170.
(R,K){20,20}
(K){20}
47: RAVAQ KKKKKKKKKKKKKKKKKKKR KKKGG
(K){20}
48: AVAQ KKKKKKKKKKKKKKKKKKKR KKKGG
(K){20}
49: VAQK KKKKKKKKKKKKKKKKKKKR KGGGV
(K){20}
50: AQKK KKKKKKKKKKKKKKKKKKKR GGGVL

AAO03284 ck: 9650 len: 115 i Aao03284 Human polypeptide SEQ ID NO 17176.
(R,K){20,20}
(K){20}
22: KALVS KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
23: ALVSK KKKKKKKKKKKKKKKKKKKR KKKK
(K){20}
24: LVSKK KKKKKKKKKKKKKKKKKKKR KKKR
(K){20}
25: VSKK KKKKKKKKKKKKKKKKKKKR KKRGG
(K){20}
26: SKKK KKKKKKKKKKKKKKKKKKKR KKRGG
(K){20}
27: KKKK KKKKKKKKKKKKKKKKKKKR KRGA

28: KKKK (K) {20} RGGL
(R, K) {20}
29: KKKK KKKKKKKKKKKKKKKKKR GGLK
AAO03288 ck: 1109 len: 117 ! Aao03288 Human polypeptide SEQ ID NO 17180.
1
(R, K) {20, 20}
31: SCSYP KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
32: CSYPK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
33: SYPK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
34: YPKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
35: PKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
36: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
37: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
38: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
39: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
40: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
41: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
42: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
43: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
44: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
45: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
46: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
47: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
48: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
49: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
50: KKKK KKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
51: KKKK KKKKKKKKKKKKKKKKKK KKKK

52: KKKK (R, K) {20} KTXG
(R, K) {20}
53: KKKK KKKKKKKKKKKKKKKKKK TXGG
AAO03375 ck: 693 len: 54 ! Aao03375 Human polypeptide SEQ ID NO 17267.
1
(R, K) {20, 20}
25: TYLB KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
26: XLBK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
27: LBKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
28: BEKK KKKKKKKKKKKKKKKKKK KKKK
AAO03459 ck: 5217 len: 58 ! Aao03459 Human polypeptide SEQ ID NO 17351.
1
(R, K) {20, 20}
28: XMCVF KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
29: MCVFK KKKKKKKKKKKKKKKKKK KKKK
AAO03475 ck: 315 len: 138 ! Aao03475 Human polypeptide SEQ ID NO 17367.
1
(R, K) {20, 20}
44: LXLTK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
45: LXLTK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
46: XLTK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
47: LTKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
48: TKKK KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
49: KKKK KKKKKKKKKKKKKKKKKK KKKK
AAO03615 ck: 7613 len: 128 ! Aao03615 Human polypeptide SEQ ID NO 17507.
1
(R, K) {20, 20}
28: PLPRX KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
29: LPRX KKKKKKKKKKKKKKKKKK KKKK
(K) {20}
30: PRXX KKKKKKKKKKKKKKKKKK KKKK
AAO03623 ck: 3884 len: 116 ! Aao03623 Human polypeptide SEQ ID NO 17515.
1
(R, K) {20, 20}
77: TPSRA KKKKKKKKKKKKKKKKKK KKKK

78: PSRAK (K) {20}
79: SPRAK (K) {20}
80: RAKKK (K) {20}
81: AKKKK (K) {20}
82: KKKKK (K) {20}
83: KKKKK (K) {20}
84: KKKKK (K) {20}
85: KKKKK (K) {20}
86: KKKKK (K) {20}
87: KKKKK (K) {20}
88: KKKKK (K) {20}

1
AAO03700 ck: 1279 len: 43 1 Aao03700 Human polypeptide SEQ ID NO 17592.
(R,K) {20,20}
(K) {20}
19: YSQRL KKKKKKKKKKKKKKKKKKKKK GGGPF

1
AAO03703 ck: 7016 len: 113 1 Aao03703 Human polypeptide SEQ ID NO 17595.
(R,K) {20,20}
(K) {20}
58: IMDAE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: WDAEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
60: DAEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
61: AEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
62: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KSPGG
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK SPGGA

1
AAO03766 ck: 8808 len: 81 1 Aao03766 Human polypeptide SEQ ID NO 17658.
(R,K) {20,20}

49: TTAC (K) {20}
50: TTACK (K) {20}
51: TACKK (K) {20}
52: ACKKK (K) {20}
53: CKKKK (K) {20}
54: KKKKK (K) {20}
55: KKKKK (K) {20}
56: KKKKK (K) {20}
57: KKKKK (K) {20}

1
AAO03841 ck: 8734 len: 100 1 Aao03841 Human polypeptide SEQ ID NO 17733.
(R,K) {20,20}
(K) {20}
36: KQHYP KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

37: QHYPK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: HYPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: YPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
40: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KRTKQ
(K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK RTKOK
(R,K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKK TKOKK

1
AAO03906 ck: 4312 len: 100 1 Aao03906 Human polypeptide SEQ ID NO 17798.
(R,K) {20,20}
(K) {20}
34: NKONQ KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: KONOK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: QNOKK KKKKKKKKKKKKKKKKKKKKK KKKKK

37: NQKKK (K) {20}
 38: QKKKK (K) {20}
 39: KKKKK (K) {20}
 40: KKKKK (K) {20}
 41: KKKKK (K) {20}
 42: KKKKK (K) {20}
 43: KKKKK (K) {20}
 44: KKKKK (K) {20}
 45: KKKKK (K) {20}
 46: KKKKK (K) {20}

AA003967 ck: 7917 len: 53 ! Aa003967 Human polypeptide SEQ ID NO 17859.

16: KKKKK (R,K) {20,20}
 17: LKSKK (K) {20}
 18: XKSKK (K) {20}
 19: KSKKK (K) {20}
 20: SKKKK (K) {20}
 21: KKKKK (K) {20}
 22: KKKKK (K) {20}
 23: KKKKK (K) {20}
 24: KKKKK (K) {20}
 25: KKKKK (K) {20}

AA004512 ck: 4277 len: 62 ! Aa004512 Human polypeptide SEQ ID NO 18404.

(R,K) {20,20}
 24: KKKKK (K) {20}

13: LITPL (R,K) {20,20}
 14: ITPLK (K) {20}
 15: TPLKK (K) {20}
 16: FLKKK (K) {20}
 17: LKKKK (K) {20}
 18: KKKKK (K) {20}
 19: KKKKK (K) {20}

AA004644 ck: 2038 len: 70 ! Aa004644 Human polypeptide SEQ ID NO 18536.

23: AKPPT (R,K) {20,20}

AA004645 ck: 7825 len: 118 ! Aa004645 Human polypeptide SEQ ID NO 18537.

46: IKSPL (R,K) {20,20}
 47: KSFLK (K) {20}
 48: SPLRK (K) {20}
 49: FLRKK (K) {20}
 50: LRKKK (K) {20}
 51: RKKKK (K) {20}
 52: KKKKK (K) {20}
 53: KKKKK (K) {20}
 54: KKKKK (K) {20}
 55: KKKKK (K) {20}
 56: KKKKK (K) {20}
 57: KKKKK (K) {20}
 58: KKKKK (K) {20}

59: KKKKK (K) {20} KKKPR
60: KKKKK (K) {20} KKKPRG
61: KKKKK (K) {20} KKKPRG
62: KKKKK (K) {20} KKKPRG

AA004647 ck: 4805 len: 58 ! Aa004647 Human polypeptide SEQ ID NO 18539.

(R,K) {20,20}
28: KPTRP KKKKK (K) {20} KKKKK
29: PTPRP KKKKK (K) {20} KKKKK
30: TRPRK KKKKK (K) {20} KKKKK
31: RPRKK KKKKK (K) {20} KKKKK
32: PRKKK KKKKK (K) {20} KKKKK
33: KKKKK (K) {20} KKKKK
34: KKKKK (K) {20} KKKKK
35: KKKKK (K) {20} KKKKK
36: KKKKK (K) {20} KKKKK
37: KKKKK (K) {20} KKKKK
38: KKKKK (K) {20} KKKKK S

AA004674 ck: 2036 len: 33 ! Aa004674 Human polypeptide SEQ ID NO 18566.

(R,K) {20,20}
9: WCYIT KKKKK (K) {20} KKKRG
10: CYITK KKKKK (K) {20} KKKRG
11: YITKK KKKKK (K) {20} KKKRG

AA004679 ck: 7190 len: 31 ! Aa004679 Human polypeptide SEQ ID NO 18571.

(R,K) {20,20}
8: CWFTQ KKKKK (K) {20} KKKRG
9: WFTQK KKKKK (K) {20} KKKRG

AA004682 ck: 5605 len: 60 ! Aa004682 Human polypeptide SEQ ID NO 18574.

(R,K) {20,20}
2: L KKKKK (K) {20} KKKKK
3: LK KKKKK (K) {20} KKKKK
4: LKK KKKKK (K) {20} KKKKK
5: LKKK KKKKK (K) {20} KKKKK

AA004690 ck: 7157 len: 81 ! Aa004690 Human polypeptide SEQ ID NO 18582.

(R,K) {20,20}
1: KKKKK (K) {20} KKKKK
2: K KKKKK (K) {20} KKKKK
3: KK KKKKK (K) {20} KKKKK
4: KKK KKKKK (K) {20} KKKKK
5: KKKK KKKKK (K) {20} KKKKK
6: KKKKK (K) {20} KKKKK
7: KKKKK (K) {20} KKKKK
8: KKKKK (K) {20} KKKKK

AA004715 ck: 6984 len: 35 ! Aa004715 Human polypeptide SEQ ID NO 18607.

(R,K) {20,20}
8: LGSKD KKKKK (K) {20} KKKKK
9: GSKDK KKKKK (K) {20} KKKKK
10: SKDKK KKKKK (K) {20} KKKKK

AA004743 ck: 9412 len: 54 ! Aa004743 Human polypeptide SEQ ID NO 18635.

(R,K) {20,20}
21: KINKL KKKKK (K) {20} KKKKK
22: INKLK KKKKK (K) {20} KKKKK
23: NKLKK KKKKK (K) {20} KKKKK

AA004747 ck: 8399 len: 39 ! Aa004747 Human polypeptide SEQ ID NO 18639.

1

(R,K){20,20}
13: RTGFV KKKKKKKKKKKKKKKKKKK KKRGG
(K){20}
14: TGFVK KKKKKKKKKKKKKKKKKKK KRGGG
(K){20}
15: GFVKK KKKKKKKKKKKKKKKKKKK RGGGF
(R,K){20}
16: FVKKK KKKKKKKKKKKKKKKKKKK GGGF

AA004752 ck: 1021 len: 71 ! Aa004752 Human polypeptide SEQ ID NO 18644.

1

(R,K){20,20}
19: OEOGL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: EOGLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: OGLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: GLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKGGG
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KGGGL
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK GGGLL

AA004755 ck: 5521 len: 59 ! Aa004755 Human polypeptide SEQ ID NO 18647.

1

(R,K){20,20}
31: INSLE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: NSLEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: SLEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: LEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
35: EKKKK KKKKKKKKKKKKKKKKKKK KKKKT
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKTT
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKT
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KT
(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKK T

AA004756 ck: 6952 len: 26 ! Aa004756 Human polypeptide SEQ ID NO 18648.

1

(R,K){20,20}
4: FFY KKKKKKKKKKKKKKKKKKK SSS
(K){20}

AA004758 ck: 7836 len: 115 ! Aa004758 Human polypeptide SEQ ID NO 18650.

1

(R,K){20,20}
7: PFYQL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: FYQLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: YQLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: QLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

27: KKKK (K) {20}
28: KKKK (K) {20}
29: KKKK (K) {20}
30: KKKK (K) {20}
31: KKKK (K) {20}
32: KKKK (K) {20}
33: KKKK (K) {20}
34: KKKK (K) {20}
35: KKKK (K) {20}
36: KKKK (K) {20}
37: KKKK (K) {20}
38: KKKK (K) {20}
39: KKKK (K) {20}
40: KKKK (K) {20}
41: KKKK (K) {20}
42: KKKK (K) {20}
43: KKKK (K) {20}
44: KKKK (K) {20}
45: KKKK (K) {20}
46: KKKK (K) {20}
47: KKKK (K) {20}
48: KKKK (K) {20}
49: KKKK (K) {20}
50: KKKK (K) {20}

1
AAO04764 ck: 2035 len: 54 ! Aao04764 Human polypeptide SEQ ID NO 18656.
9: GDSSL (R,K) {20,20}
10: DSSLK (K) {20}
11: SSLK (K) {20}
12: SLKK (K) {20}
13: LKKK (K) {20}
14: KKKK (K) {20}
1
AAO04802 ck: 1223 len: 105 ! Aao04802 Human polypeptide SEQ ID NO 18694.
34: FFSRQ (R,K) {20,20}
AAO04854 ck: 4121 len: 58 ! Aao04854 Human polypeptide SEQ ID NO 18746.
13: TPFRK (R,K) {20,20}
14: PFRK (K) {20}
15: FRKK (K) {20}
16: RAKK (K) {20}
17: AKKK (K) {20}
18: KKKK (K) {20}
19: KKKK (K) {20}
20: KKKK (K) {20}
21: KKKK (K) {20}
22: KKKK (K) {20}
23: KKKK (K) {20}
24: KKKK (K) {20}
25: KKKK (K) {20}
26: KKKK (K) {20}

27: KKKKK (K) {20} KKKKK
 28: KKKKK (K) {20} KKKKK
 29: KKKKK (K) {20} KKKKK
 30: KKKKK (K) {20} KKKKK
 31: KKKKK (K) {20} KKKKK
 32: KKKKK (R,K) {20} KKKKK

AA004856 ck: 4861 len: 135 ! Aa004856 Human polypeptide SEQ ID NO 18748.

41: LTLTT (R,K) {20,20} KKKKK
 42: TLTTK (K) {20} KKKKK
 43: LTTKK (K) {20} KKKKK
 44: TTYKK (K) {20} KKKKK
 45: TKKKK (K) {20} KKKKK
 46: KKKKK (K) {20} KKKKK
 47: KKKKK (K) {20} KKKKK
 48: KKKKK (K) {20} KKKKK
 49: KKKKK (K) {20} KKKKK
 50: KKKKK (K) {20} KKKKK
 51: KKKKK (K) {20} KKKKK
 52: KKKKK (K) {20} KKKKK
 53: KKKKK (K) {20} KKKKK
 AA004872 ck: 3224 len: 93 ! Aa004872 Human polypeptide SEQ ID NO 18764.
 8: FLVYL (R,K) {20,20} KKKKK
 9: LVYLL (K) {20} KKKKK

10: YKLKK (K) {20} KKKKK
 11: KLKKK (K) {20} KKKKK
 12: LKKKK (R,K) {20} KKKKK

AA004873 ck: 7719 len: 66 ! Aa004873 Human polypeptide SEQ ID NO 18765.

21: SFLIE (R,K) {20,20} KKKKK
 22: FLIEK (K) {20} KKKKK

AA004874 ck: 5753 len: 58 ! Aa004874 Human polypeptide SEQ ID NO 18766.

10: SEKLP (R,K) {20,20} KKKKK
 11: EKLPK (K) {20} KKKKK
 12: KLPKK (K) {20} KKKKK

AA004881 ck: 8841 len: 115 ! Aa004881 Human polypeptide SEQ ID NO 18773.

24: ITPHP (R,K) {20,20} KKKKK
 25: TPHPK (K) {20} KKKKK
 26: PPHPK (K) {20} KKKKK
 27: HPKKK (K) {20} KKKKK
 28: PKKKK (K) {20} KKKKK
 29: KKKKK (K) {20} KKKKK
 30: KKKKK (K) {20} KKKKK
 31: KKKKK (K) {20} KKKKK
 AA004917 ck: 3999 len: 103 ! Aa004917 Human polypeptide SEQ ID NO 18809.
 1: KKKKK (R,K) {20,20} KKKKK

AA004928 ck: 23 len: 43 ! Aa004928 Human polypeptide SEQ ID NO 18820.
 19: FKKEK (R,K) {20,20} KKKKK

1

AA004969 ck: 4345 len: 57 ! Aa004969 Human polypeptide SEQ ID NO 18861.

(R,K){20,20}

(K){20}

26: LMGTG KKKKKKKKKKKKKKKKKKKKK

(K){20}

27: MGTSG KKKKKKKKKKKKKKKKKKKKK

(K){20}

28: GTSRK KKKKKKKKKKKKKKKKKKKKK

(K){20}

29: TSKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

30: SKRKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

33: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

34: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKK

AA005000 ck: 5847 len: 38 ! Aa005000 Human polypeptide SEQ ID NO 18892.

(R,K){20,20}

(K){20}

5: SGRS KKKKKKKKKKKKKKKKKKKKK

(K){20}

6: SGRSK KKKKKKKKKKKKKKKKKKKKK

(K){20}

7: GRSKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

8: RSKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

9: SKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKK

14: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKK

(R,K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKK

AA005081 ck: 2260 len: 40 ! Aa005081 Human polypeptide SEQ ID NO 18973.

(R,K){20,20}

(K){20}

9: SPASP KKKKKKKKKKKKKKKKKKKKK

(K){20}

10: RASPK KKKKKKKKKKKKKKKKKKKKK

(K){20}

11: ASPKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

12: SPKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

13: PKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKK

AA005130 ck: 6597 len: 116 ! Aa005130 Human polypeptide SEQ ID NO 19022.

(R,K){20,20}

(K){20}

14: ETPFQ KKKKKKKKKKKKKKKKKKKKK

(K){20}

15: TPFQK KKKKKKKKKKKKKKKKKKKKK

(R,K){20}

16: PFQKK KKKKKKKKKKKKKKKKKKKKK

AA005191 ck: 4399 len: 85 ! Aa005191 Human polypeptide SEQ ID NO 19083.

(R,K){20,20}

(R,K){20}

30: GYIQL KKKKKKKKKKKKKKKKKKKKK

AA005260 ck: 3618 len: 55 ! Aa005260 Human polypeptide SEQ ID NO 19152.

(R,K){20,20}

(K){20}

15: PHHQK KKKKKKKKKKKKKKKKKKKKK

(K){20}

16: HGHQK KKKKKKKKKKKKKKKKKKKKK

17: KHQKK (K) {20}
18: HQKK (K) {20}
19: QKKK (K) {20}
20: KKKK (K) {20}
21: KKKK (K) {20}
22: KKKK (K) {20}
23: KKKK (K) {20}

AA005369 ck: 1417 len: 70 1 Aa005369 Human polypeptide SEQ ID NO 19261.

1

(R,K) {20,20}
(K) {20}

22: KYHET KKKKKKKKKKKKKKKKKKKKKKKKKKK
23: YHETK KKKKKKKKKKKKKKKKKKKKKKKKKKK

24: HETKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

25: ETKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

26: TKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

31: KKKKK (R,K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

AA005372 ck: 9880 len: 43 1 Aa005372 Human polypeptide SEQ ID NO 19264.

1

(R,K) {20,20}
(K) {20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

6: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

7: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

8: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

9: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

10: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

11: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

12: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

13: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

14: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

15: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

16: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

17: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AA005384 ck: 715 len: 23 1 Aa005384 Human polypeptide SEQ ID NO 19276.

(R,K) {20,20}
(K) {20}

4: LFS KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AA005499 ck: 7813 len: 76 1 Aa005499 Human polypeptide SEQ ID NO 19391.

(R,K) {20,20}
(K) {20}

7: VQPQQ KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

8: QPQQK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

9: PQQKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

10: QQQKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

11: QKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

12: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

13: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

14: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

15: KKKKK (K) (20) KKKKK
16: KKKKK (K) (20) KKKKK
17: KKKKK (K) (20) KKKKK
18: KKKKK (K) (20) KKKKK
19: KKKKK (K) (20) KKKKK
20: KKKKK (K) (20) KKKKK
21: KKKKK (K) (20) KKKKK
22: KKKKK (K) (20) KKKKK
23: KKKKK (K) (20) KKKKK
24: KKKKK (K) (20) KKKKK
25: KKKKK (K) (20) KKKKK

AA005500 ck: 7443 len: 53 1 Aa005500 Human polypeptide SEQ ID NO 19392.

1 (R, K) (20, 20)
(K) (20)

5: QPOS KKKKKKKKKKKKKKKKKKK KASQ

6: QPOS KKKKKKKKKKKKKKKKKKK KASQK

7: POSK KKKKKKKKKKKKKKKKKKK ASQK

AA005530 ck: 7354 len: 75 1 Aa005530 Human polypeptide SEQ ID NO 19422.

1 (R, K) (20, 20)
(K) (20)

11: TTYFL KKKKKKKKKKKKKKKKKKK KKKKK

12: TYFLK KKKKKKKKKKKKKKKKKKK KKKKK

13: YPLK KKKKKKKKKKKKKKKKKKK KKKKK

14: FLKK KKKKKKKKKKKKKKKKKKK KKKKK

15: LKKK KKKKKKKKKKKKKKKKKKK KKKKK

16: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

17: KKKK (K) (20) KKKKK

18: KKKK (K) (20) KKKKK

19: KKKKK (K) (20) KKKKK
20: KKKKK (K) (20) KKKKK
21: KKKKK (K) (20) KKKKK
22: KKKKK (K) (20) KKKKK
23: KKKKK (K) (20) KKKKK
24: KKKKK (K) (20) KKKKK
25: KKKKK (K) (20) KKKKK
26: KKKKK (K) (20) KKKKK
27: KKKKK (K) (20) KKKKK
28: KKKKK (K) (20) KKKKK
29: KKKKK (K) (20) KKKKK
30: KKKKK (K) (20) KKKKK
31: KKKKK (K) (20) KKKKK
32: KKKKK (K) (20) KKKKK
33: KKKKK (K) (20) KKKKK
34: KKKKK (K) (20) KKKKK
35: KKKKK (K) (20) KKKKK
36: KKKKK (K) (20) KKKKK
37: KKKKK (K) (20) KKKKK
38: KKKKK (K) (20) KKKKK
39: KKKKK (K) (20) KKKKK
40: KKKKK (K) (20) KKKKK
41: KKKKK (K) (20) KKKKK
42: KKKKK (K) (20) KKKKK

1 AA005665 ck: 432 len: 28 ! Aa005665 Human polypeptide SEQ ID NO 19557.
(R,K){20,20}
(K){20}
6: DFLIQ KKKKKKKKKKKKKKKKKKK RGG
(R,K){20}
7: FLIQK KKKKKKKKKKKKKKKKKR GG

1 AA006186 ck: 998 len: 88 ! Aa006186 Human polypeptide SEQ ID NO 20078.
(R,K){20,20}
(K){20}
28: SLIPK KKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
29: LIPK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
30: LIPK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
31: PKKK KKKKKKKKKKKKKKKKKKK KGGGG
(K){20}
32: XKKK KKKKKKKKKKKKKKKKKKK GGGGF

1 AA006357 ck: 4679 len: 52 ! Aa006357 Human polypeptide SEQ ID NO 20249.
(R,K){20,20}
(K){20}
9: FTCLI KKKKKKKKKKKKKKKKKKK KKKIK
(K){20}
10: TCLI KKKKKKKKKKKKKKKKKKK KKIKX
(K){20}
11: CLIK KKKKKKKKKKKKKKKKKKK KIKKK
(K){20}
12: LIKK KKKKKKKKKKKKKKKKKKK IKKKK

1 AA006429 ck: 2585 len: 71 ! Aa006429 Human polypeptide SEQ ID NO 20321.
(R,K){20,20}
(K){20}
29: VIIMK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: IIMK KKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
31: IMKK KKKKKKKKKKKKKKKKKKK KKKGR
(K){20}
32: MKKK KKKKKKKKKKKKKKKKKKK KKGKG
(K){20}
33: XKKK KKKKKKKKKKKKKKKKKKK KKGKG
(K){20}
34: KKKK KKKKKKKKKKKKKKKKKKK KRGGA

1 AA006922 ck: 7296 len: 111 ! Aa006922 Human polypeptide SEQ ID NO 20814.
(R,K){20,20}
(K){20}
13: HSEL KKKKKKKKKKKKKKKKKKK GGGPK

1 AA007241 ck: 3134 len: 40 ! Aa007241 Human polypeptide SEQ ID NO 21133.
(R,K){20,20}
(K){20}
4: YFP KKKKKKKKKKKKKKKKKKK RKKKX
(R,K){20}
5: YFP KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
6: YFPK KKKKKKKKKKKKKKKKKKK KKKKG
(R,K){20}
7: FPKK KKKKKKKKKKKKKKKKKKK KKKGG
(R,K){20}
8: PKKK KKKKKKKKKKKKKKKKKKK XKKGG

1 AA007354 ck: 6513 len: 133 ! Aa007354 Human polypeptide SEQ ID NO 21246.
(R,K){20,20}
(K){20}
36: KYHL KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
37: XYLK KKKKKKKKKKKKKKKKKKK KKKGP
(K){20}
38: YHLK KKKKKKKKKKKKKKKKKKK KGGPL
(K){20}
39: HLKK KKKKKKKKKKKKKKKKKKK GGPKL

1 AA007410 ck: 693 len: 80 ! Aa007410 Human polypeptide SEQ ID NO 21302.
(R,K){20,20}
(K){20}
56: HNPI KKKKKKKKKKKKKKKKKKK RGGGF
(R,K){20}
57: KPIK KKKKKKKKKKKKKKKKKKK GGGF

1 AA007412 ck: 4623 len: 166 ! Aa007412 Human polypeptide SEQ ID NO 21304.
(R,K){20,20}
(K){20}
93: VNTQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
94: NTQK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
95: TQKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
96: QKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
97: XKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
98: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
99: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
100: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
107: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AA007505 ck: 3528 len: 72 ! Aa007505 Human polypeptide SEQ ID NO 21397.

1

(R, K) {20, 20}
(K) {20}

14: HLVEA KKKKKKKKKKKKKKKKKKK KKKKK

15: LVEAK KKKKKKKKKKKKKKKKKKK KKKKK

16: VEAkk KKKKKKKKKKKKKKKKKKK KKKKK

17: EAkkK KKKKKKKKKKKKKKKKKKK KKKKK

18: AKkkK KKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AA007509 ck: 4998 len: 140 ! Aa007509 Human polypeptide SEQ ID NO 21401.

1

(R, K) {20, 20}
(K) {20}

17: KVVKE KKKKKKKKKKKKKKKKKKK KKKKK

18: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

20: XKKKK KKKKKKKKKKKKKKKKKKK KKKKK

21: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AA007594 ck: 7271 len: 93 ! Aa007594 Human polypeptide SEQ ID NO 21486.

1

(R, K) {20, 20}
(K) {20}

9: FLLGG KKKKKKKKKKKKKKKKKKK KKKKK

10: LLLGG KKKKKKKKKKKKKKKKKKK KKKKK

11: LLGGK KKKKKKKKKKKKKKKKKKK KKKKK

12: LGKKK KKKKKKKKKKKKKKKKKKK KKKKK

13: GKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AA007607 ck: 2434 len: 72 ! Aa007607 Human polypeptide SEQ ID NO 21499.

1

(R, K) {20, 20}
(K) {20}

21: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

48: GLFXE KKKKKKKKKKKKKKKKKKK KKKKK

AA007610 ck: 5270 len: 74 ! Aa007610 Human polypeptide SEQ ID NO 21502.

1

(R, K) {20, 20}
(K) {20}

19: TELTI KKKKKKKKKKKKKKKKKKK KKKKK

20: ELTIK KKKKKKKKKKKKKKKKKKK KKKKK

21: LTIKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKK
22: TIRKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKG
23: IKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKG
24: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKGCG
25: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KGGGA
26: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} GGGAL

AA007755 ck: 6069 len: 90 ! Aa007755 Human polypeptide SEQ ID NO 21647.

(R,K) {20,20}
35: KNOSS KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
36: NQSSK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
37: QSSKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
38: SSKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKG
39: SKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKG
40: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKGGP
41: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KGGPL
42: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} GGPPL

AA007762 ck: 5115 len: 41 ! Aa007762 Human polypeptide SEQ ID NO 21654.

(R,K) {20,20}
18: IPSLK KKKKKKKKKKKKKKKKKKK (K) {20} KKKK
19: PSLKK KKKKKKKKKKKKKKKKKKK (K) {20} KKK
20: SLKKK KKKKKKKKKKKKKKKKKKK (K) {20} KK
21: LKKKK KKKKKKKKKKKKKKKKKKK (K) {20} K
22: XKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK

AA007775 ck: 9827 len: 90 ! Aa007775 Human polypeptide SEQ ID NO 21667.

(R,K) {20,20}
46: PHKPE KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK

47: HXPEK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKG
48: XPEKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKG
49: PEKKK KKKKKKKKKKKKKKKKKKK (K) {20} KGGGA
50: EKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KGGAP
51: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} GGAPK

AA007806 ck: 1959 len: 99 ! Aa007806 Human polypeptide SEQ ID NO 21698.

(R,K) {20,20}
41: KLHCI KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
42: LHCIC KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
43: HCICK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
44: CIKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKL
45: IKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKLW
46: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KKLWG
47: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} KLWGG
48: KKKKK KKKKKKKKKKKKKKKKKKK (K) {20} LMGGG

AA007872 ck: 8003 len: 135 ! Aa007872 Human polypeptide SEQ ID NO 21764.

(R,K) {20,20}
52: TRKKQ KKKKKKKKKKKKKKKKKKK (K) {20} GGGLL

AA007874 ck: 8007 len: 111 ! Aa007874 Human polypeptide SEQ ID NO 21766.

(R,K) {20,20}
90: EEPIS KKKKKKKKKKKKKKKKKKK (K) {20} KG
91: EPSIK KKKKKKKKKKKKKKKKKKK (K) {20} G

AA007876 ck: 8130 len: 66 ! Aa007876 Human polypeptide SEQ ID NO 21768.

(R,K) {20,20}
17: IYSFE KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
18: YSFEK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
19: SFEKK KKKKKKKKKKKKKKKKKKK (K) {20} KKKKQ

20: PEKKK (K) {20}
21: EKKKK (K) {20}
22: KKKKK (K) {20}
23: KKKKK (K) {20}

AA007976 ck: 6471 len: 97 ! Aa007976 Human polypeptide SEQ ID NO 21868.

43: PPIST (R,K) {20,20}
44: PISTK (K) {20}
45: ISTKK (K) {20}
46: STKKK (K) {20}
47: TKKKK (K) {20}
48: KKKKK (K) {20}
49: KKKKK (K) {20}
50: KKKKK (K) {20}
51: KKKKK (K) {20}
52: KKKKK (K) {20}
53: KKKKK (K) {20}
54: KKKKK (K) {20}
55: KKKKK (K) {20}

AA007981 ck: 9551 len: 81 ! Aa007981 Human polypeptide SEQ ID NO 21873.

54: KPCIY (R,K) {20,20}
55: PCIYK (K) {20}
56: CIYKK (K) {20}
57: IYKKK (K) {20}
58: YKKKK (K) {20}

59: KKKKK (K) {20}
60: KKKKK (K) {20}
61: KKKKK (K) {20}
62: KKKKK (K) {20}

AA008018 ck: 9768 len: 48 ! Aa008018 Human polypeptide SEQ ID NO 21910.

24: KISQV (R,K) {20,20}
25: ISQVR (K) {20}
26: SQVRK (K) {20}
27: QVRKK (K) {20}
28: VRKKK (K) {20}
29: RKKKK (K) {20}

AA008121 ck: 5381 len: 86 ! Aa008121 Human polypeptide SEQ ID NO 22013.

20: HFGLL (R,K) {20,20}
20: HFGLL (K) {20}

AA008232 ck: 782 len: 104 ! Aa008232 Human polypeptide SEQ ID NO 22124.

49: ISQON (R,K) {20,20}
49: ISQON (K) {20}
50: SQONK (K) {20}

AA008450 ck: 6228 len: 31 ! Aa008450 Human polypeptide SEQ ID NO 22342.

5: NNSA (R,K) {20,20}
6: NNSAK (K) {20}
7: NSAKK (K) {20}

AA008458 ck: 1813 len: 59 ! Aa008458 Human polypeptide SEQ ID NO 22350.

8: LGWBE (R,K) {20,20}
8: LGWBE (K) {20}

9: GWEK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
10: GWEK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
11: WEKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
12: EKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
13: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAO08460 ck: 5334 len: 62 ! Aao08460 Human polypeptide SEQ ID NO 22352.

1

(R,K) {20,20}
(K) {20}

22: KEFST KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

23: EFSTK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

24: FSTK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

25: STKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

26: TKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

27: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

28: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

29: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

30: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

31: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

32: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

33: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

34: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAO08466 ck: 8728 len: 107 ! Aao08466 Human polypeptide SEQ ID NO 22358.

1

(R,K) {20,20}
(K) {20}

57: FMGV KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

58: WGVK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

59: GVKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

60: GVKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

61: VKKK (K) {20}
KKKKKKKKKKKKKKKKKKKKKKKKKKRRRG

62: KKKK (K) {20}
KKKKKKKKKKKKKKKKKKKKKKKKKKRRKR

63: KKKK (R,K) {20}
KKKKKKKKKKKKKKKKKKKKKKKKKKRGRK

64: KKKK (R,K) {20}
KKKKKKKKKKKKKKKKKKKKKKKKKKRGRK

65: KKKK (R,K) {20}
KKKKKKKKKKKKKKKKKKKKKKKKKKRGRK

AAO08580 ck: 5164 len: 62 ! Aao08580 Human polypeptide SEQ ID NO 22472.

1

(R,K) {20,20}
(K) {20}

30: PILL KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAO08591 ck: 900 len: 46 ! Aao08591 Human polypeptide SEQ ID NO 22483.

1

(R,K) {20,20}
(K) {20}

14: LLSQ KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAO08609 ck: 5151 len: 74 ! Aao08609 Human polypeptide SEQ ID NO 22501.

1

(R,K) {20,20}
(K) {20}

18: KCVL KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

19: CVLK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAO08612 ck: 557 len: 46 ! Aao08612 Human polypeptide SEQ ID NO 22504.

1

(R,K) {20,20}
(K) {20}

20: TFCIM KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

21: FCIMK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

22: CIMK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

23: IMKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAO08616 ck: 4555 len: 41 ! Aao08616 Human polypeptide SEQ ID NO 22508.

1

(R,K) {20,20}
(K) {20}

6: HCLP KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

7: CALPK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

8: ALPK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

9: LPKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

10: PKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

11: KKKKK (K) {20} KKKKK
12: KKKKK (K) {20} KKKKK
13: KKKKK (K) {20} KKKKK
14: KKKKK (K) {20} KKKKK
15: KKKKK (K) {20} KKKKK
16: KKKKK (K) {20} KKKKK
17: KKKKK (K) {20} KKKKK
18: KKKKK (K) {20} KKKKK
19: KKKKK (K) {20} KKKKK
20: KKKKK (K) {20} KKKKK
21: KKKKK (K) {20} K
22: KKKKK (K) {20} KKKKK

AA008623 ck: 9243 len: 119 ! Aa008623 Human polypeptide SEQ ID NO 22515.
(R,K) {20,20}
(R,K) {20}

21: DSKQE KKKKKKKKKKKKKKKKKR GGAFF

AA008624 ck: 5590 len: 65 ! Aa008624 Human polypeptide SEQ ID NO 22516.
(R,K) {20,20}
(K) {20}

11: EGNMW KKKKKKKKKKKKKKKKKK KKKKK
12: GNNWK KKKKKKKKKKKKKKKKKK KKKKK
13: NNWKK (K) {20} KKKKK
14: NNWKK (K) {20} KKKKKKKKKKKKKKKKK KKKKK
15: WKKKK (K) {20} KKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (K) {20} KKKKKKKKKKKKKKKKK KKKKK

AA008631 ck: 6430 len: 135 ! Aa008631 Human polypeptide SEQ ID NO 22523.
(R,K) {20,20}
(K) {20}

30: TPSRA KKKKKKKKKKKKKKKKKKK KKKKK

31: PSRAK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
32: SPRAK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
33: PAKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
34: AKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
35: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
36: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
37: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
38: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
39: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
40: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
41: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
42: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
43: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
44: KKKKK (R,K) {20} KKKKKKKKKKKKKKKKKKK KKKKK

AA008644 ck: 3202 len: 80 ! Aa008644 Human polypeptide SEQ ID NO 22536.
(R,K) {20,20}
(K) {20}

2: E KKKKKKKKKKKKKKKKKKKKK KKKKK
3: EK KKKKKKKKKKKKKKKKKKKKK KKKKK
4: EKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
5: EKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
6: EKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK

11: KKKK (K) {20}
12: KKKK (K) {20}
13: KKKK (K) {20}
14: KKKK (K) {20}
15: KKKK (K) {20}
16: KKKK (K) {20}
17: KKKK (K) {20}
18: KKKK (K) {20}
19: KKKK (K) {20}

AA008653 ck: 6448 len: 63 1 Aa008653 Human polypeptide SEQ ID NO 22545.

12: PIKLT (R,K) {20,20}
13: IKLT (K) {20}
14: KLT (K) {20}
15: LT (K) {20}
16: TK (K) {20}
17: K (K) {20}
18: K (K) {20}
19: K (K) {20}
20: K (R,K) {20}

AA008686 ck: 6936 len: 81 1 Aa008686 Human polypeptide SEQ ID NO 22578.

23: HPVNH (R,K) {20,20}
24: PVNH (K) {20}
25: VNH (K) {20}
26: NH (K) {20}

27: HKKK (K) {20}
28: KKKK (K) {20}
29: KKKK (K) {20}
30: KKKK (K) {20}
31: KKKK (K) {20}
32: KKKK (K) {20}
33: KKKK (K) {20}
34: KKKK (K) {20}
35: KKKK (K) {20}
36: KKKK (K) {20}

37: KKKK (K) {20}

AA008701 ck: 588 len: 46 1 Aa008701 Human polypeptide SEQ ID NO 22593.

7: INANS (R,K) {20,20}
8: NANS (K) {20}
9: ANS (K) {20}
10: NS (K) {20}
11: S (K) {20}
12: K (K) {20}
13: K (K) {20}

AA008707 ck: 9359 len: 48 1 Aa008707 Human polypeptide SEQ ID NO 22599.

13: TAVTQ (R,K) {20,20}
14: AVTQ (K) {20}
15: VTQ (K) {20}

1 AA008708 ck: 9654 len: 32 ! Aa008708 Human polypeptide SEQ ID NO 22600.
(R,K){20,20}
(K){20}
8: SMTEP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: MTEPK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
10: TPEPK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
11: FPKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
12: PKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK

1 AA008717 ck: 4688 len: 99 ! Aa008717 Human polypeptide SEQ ID NO 22609.
(R,K){20,20}
(K){20}
80: SPFLI KKKKKKKKKKKKKKKKKKK

1 AA008772 ck: 1431 len: 105 ! Aa008772 Human polypeptide SEQ ID NO 22664.
(R,K){20,20}
(K){20}
23: IMNYL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: MNYLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: NYLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: YLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: LKKKK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KGGGP
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK GGGPF

1 AA008774 ck: 2529 len: 71 ! Aa008774 Human polypeptide SEQ ID NO 22666.
(R,K){20,20}
(K){20}
16: RTIKL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: TIKLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: IKLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

20: LKKKK KKKKKKKKKKKKKKKKKKK KKGCG
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKGCP
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KGGPP
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK GGGPF

1 AA008817 ck: 529 len: 28 ! Aa008817 Human polypeptide SEQ ID NO 22709.
(R,K){20,20}
(K){20}
3: AX KKKKKKKKKKKKKKKKKKK NIWGG

1 AA008820 ck: 1888 len: 84 ! Aa008820 Human polypeptide SEQ ID NO 22712.
(R,K){20,20}
(K){20}
28: CMTFS KKKKKKKKKKKKKKKKKKK RGGGF
(R,K){20}
29: MTFPS KKKKKKKKKKKKKKKKKKK GGGFI

1 AA008841 ck: 3529 len: 74 ! Aa008841 Human polypeptide SEQ ID NO 22733.
(R,K){20,20}
(K){20}
3: PQ KKKKKKKKKKKKKKKKKKK KMGCG
(K){20}
4: PQK KKKKKKKKKKKKKKKKKKK WGGGF

1 AA008857 ck: 2816 len: 49 ! Aa008857 Human polypeptide SEQ ID NO 22749.
(R,K){20,20}
(K){20}
25: PEPTS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: PPTSK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
27: PTSKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
28: TSKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
29: SKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK

1 AA008913 ck: 3132 len: 52 ! Aa008913 Human polypeptide SEQ ID NO 22805.
(R,K){20,20}
(K){20}
9: IIKSF KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: IKSFK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KSFKK KKKKKKKKKKKKKKKKKKK KKKKK

12: SPKKK (K) {20} KKKKKK
13: FKKKK (K) {20} KKKKKK
14: KKKKK (K) {20} KKKKKK
15: KKKKK (K) {20} KKKKKK
16: KKKKK (K) {20} KKKKKK
17: KKKKK (K) {20} KKKKKK
18: KKKKK (K) {20} KKKKKK
19: KKKKK (K) {20} KKKKKK
20: KKKKK (K) {20} KKKKKK
21: KKKKK (K) {20} KKKKKK
22: KKKKK (K) {20} KKKKKK
23: KKKKK (K) {20} KKKKKK
24: KKKKK (K) {20} KKKKKK
25: KKKKK (K) {20} KKKKKK
26: KKKKK (K) {20} KKKKKK
27: KKKKK (K) {20} KKKKKK
28: KKKKK (K) {20} KKKKKK
29: KKKKK (K) {20} KKKKKK
30: KKKKK (K) {20} KKKKKK
31: KKKKK (K) {20} KKKKKK
AAO08914 ck: 4641 len: 25 ! Aao08914 Human polypeptide SEQ ID NO 22806.
1 (R,K) {20,20}
6: ILMPX KKKKKKKKKKKKKKKKKKKK
AAO08943 ck: 5770 len: 75 ! Aao08943 Human polypeptide SEQ ID NO 22835.
1 (R,K) {20,20}
8: SEWAA KKKKKKKKKKKKKKKKKKKK KKKIF

9: EWAAR (K) {20} KKKKKKKKKKKKKKKKKKK KKKIF
10: WPAAR (K) {20} KKKKKKKKKKKKKKKKKKK KKKIF
11: AAKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKIF
AAO08994 ck: 2695 len: 29 ! Aao08994 Human polypeptide SEQ ID NO 22886.
1 (R,K) {20,20}
9: NLPSS (K) {20} KKKKKKKKKKKKKKKKKKK K
10: LPSHK (K) {20} KKKKKKKKKKKKKKKKKKK
AAO08995 ck: 6270 len: 26 ! Aao08995 Human polypeptide SEQ ID NO 22887.
1 (R,K) {20,20}
6: MAAPP (K) {20} KKKKKKKKKKKKKKKKKKK I
AAO09001 ck: 6400 len: 26 ! Aao09001 Human polypeptide SEQ ID NO 22893.
1 (R,K) {20,20}
5: LTSS (K) {20} KKKKKKKKKKKKKKKKKKK KK
6: LTSSK (K) {20} KKKKKKKKKKKKKKKKKKK K
7: TSSKK (K) {20} KKKKKKKKKKKKKKKKKKK
AAO09016 ck: 4378 len: 119 ! Aao09016 Human polypeptide SEQ ID NO 22908.
1 (R,K) {20,20}
10: EEEER (R,K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO09057 ck: 3712 len: 58 ! Aao09057 Human polypeptide SEQ ID NO 22949.
1 (R,K) {20,20}
5: MILN (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
6: MILNK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
7: ILNKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
8: LNKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
9: NKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO09066 ck: 2645 len: 29 ! Aao09066 Human polypeptide SEQ ID NO 22958.
1 (R,K) {20,20}

1 AAO09457 ck: 9658 len: 32 ! Aao09457 Human polypeptide SEQ ID NO 23349.
(R,K){20,20}
(K){20}
9: ALVPQ KKKKKKKKKKKKKKKKKKKKK NIKI

1 AAO09819 ck: 4709 len: 80 ! Aao09819 Human polypeptide SEQ ID NO 23711.
(R,K){20,20}
(K){20}
54: CPEVX KKKKKKKKKKKKKKKKKKKKK EKGSG

1 AAO10447 ck: 6342 len: 50 ! Aao10447 Human polypeptide SEQ ID NO 24339.
(R,K){20,20}
(K){20}
25: RPLKL KKKKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
26: PLTLK KKKKKKKKKKKKKKKKKKKKK KKKKI
(R,K){20}
27: LKLKK KKKKKKKKKKKKKKKKKKKKK KKKI
(R,K){20}
28: KLKKK KKKKKKKKKKKKKKKKKKKKK KKI
(R,K){20}
29: LKKKK KKKKKKKKKKKKKKKKKKKKK KI
(R,K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK I

1 AAO10451 ck: 3955 len: 82 ! Aao10451 Human polypeptide SEQ ID NO 24343.
(R,K){20,20}
(K){20}
59: SRASP KKKKKKKKKKKKKKKKKKKKK ARGG

1 AAO10467 ck: 7542 len: 116 ! Aao10467 Human polypeptide SEQ ID NO 24359.
(R,K){20,20}
(K){20}
57: KCEFM KKKKKKKKKKKKKKKKKKKKK KIGGG
(K){20}
58: CEFMK KKKKKKKKKKKKKKKKKKKKK IGSGA

1 AAO10564 ck: 9156 len: 98 ! Aao10564 Human polypeptide SEQ ID NO 24456.
(R,K){20,20}
(K){20}
18: NLTLT KKKKKKKKKKKKKKKKKKKKK KKEMP
(K){20}
19: LTLTK KKKKKKKKKKKKKKKKKKKKK KEMPV
(K){20}
20: LTLKK KKKKKKKKKKKKKKKKKKKKK EMPVK

1 AAO10608 ck: 9106 len: 67 ! Aao10608 Human polypeptide SEQ ID NO 24500.
(R,K){20,20}
(K){20}
42: CRUSE KKKKKKKKKKKKKKKKKKKKK KKKKD

1 43: RLSEK KKKKKKKKKKKKKKKKKKKKK KKDS
(K){20}
44: LSEKK KKKKKKKKKKKKKKKKKKKKK KKDS
(K){20}
45: SEKKK KKKKKKKKKKKKKKKKKKKKK KDS
(K){20}
46: EKKKK KKKKKKKKKKKKKKKKKKKKK DS

1 AAO10638 ck: 621 len: 56 ! Aao10638 Human polypeptide SEQ ID NO 24530.
(R,K){20,20}
(K){20}
33: CEPQP KKKKKKKKKKKKKKKKKKKKK KARG
(K){20}
34: EPQPK KKKKKKKKKKKKKKKKKKKKK ARG

1 AAO10786 ck: 7349 len: 65 ! Aao10786 Human polypeptide SEQ ID NO 24678.
(R,K){20,20}
(K){20}
24: CPXFS KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: PYFSK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: XFSKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: FSKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: SKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKV
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKVF
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKVFF
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KVVFF
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK VFFFE

1 AAO10852 ck: 5147 len: 41 ! Aao10852 Human polypeptide SEQ ID NO 24744.
(R,K){20,20}
(K){20}
20: FSRXM KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
21: SRXMK KKKKKKKKKKKKKKKKKKKKK K
(K){20}

22: RXXXXX KXXXXXXXXXXXXX

AAO10853 ck: 444 len: 98 1 Aao10853 Human polypeptide SEQ ID NO 24745.

(R,K) {20,20}

(K) {20}

14: LRAHL KXXXXXXXXXXXXX

(K) {20}

15: RAHLK KXXXXXXXXXXXXX

(K) {20}

16: AHLK KXXXXXXXXXXXXX

(K) {20}

17: HLKK KXXXXXXXXXXXXX

(K) {20}

18: LKKK KXXXXXXXXXXXXX

(K) {20}

19: KKKK KXXXXXXXXXXXXX

(K) {20}

20: KKKK KXXXXXXXXXXXXX

(K) {20}

21: KKKK KXXXXXXXXXXXXX

(K) {20}

22: KKKK KXXXXXXXXXXXXX

(K) {20}

23: KKKK KXXXXXXXXXXXXX

(K) {20}

24: KKKK KXXXXXXXXXXXXX

(K) {20}

25: KKKK KXXXXXXXXXXXXX

(K) {20}

26: KKKK KXXXXXXXXXXXXX

(K) {20}

27: KKKK KXXXXXXXXXXXXX

(K) {20}

28: KKKK KXXXXXXXXXXXXX

(R,K) {20}

29: KKKK KXXXXXXXXXXXXX

AAO10859 ck: 2690 len: 70 1 Aao10859 Human polypeptide SEQ ID NO 24751.

(R,K) {20,20}

(K) {20}

44: XGITE KXXXXXXXXXXXXX

(K) {20}

45: GITEK KXXXXXXXXXXXXX

AAO10933 ck: 849 len: 69 1 Aao10933 Human polypeptide SEQ ID NO 24825.

(R,K) {20,20}

(K) {20}

33: FINTK KXXXXXXXXXXXXX

(K) {20}

34: INTER KXXXXXXXXXXXXX

(K) {20}

35: NTEKK KXXXXXXXXXXXXX

(K) {20}

36: TEKKK KXXXXXXXXXXXXX

(K) {20}

37: EKKK KXXXXXXXXXXXXX

(K) {20}

38: KKKK KXXXXXXXXXXXXX

(K) {20}

39: KKKK KXXXXXXXXXXXXX

(K) {20}

40: KKKK KXXXXXXXXXXXXX

(K) {20}

41: KKKK KXXXXXXXXXXXXX

(K) {20}

42: KKKK KXXXXXXXXXXXXX

(K) {20}

43: KKKK KXXXXXXXXXXXXX

(K) {20}

44: KKKK KXXXXXXXXXXXXX

(K) {20}

45: KKKK KXXXXXXXXXXXXX

AAO10973 ck: 824 len: 51 1 Aao10973 Human polypeptide SEQ ID NO 24865.

(R,K) {20,20}

(R,K) {20}

16: SYFXM RXXXXXXXXXXXXX

(K) {20}

17: YFXMR KXXXXXXXXXXXXX

AAO10983 ck: 694 len: 51 1 Aao10983 Human polypeptide SEQ ID NO 24875.

(R,K) {20,20}

(K) {20}

24: HPIIX KXXXXXXXXXXXXX

(K) {20}

25: PILIX KXXXXXXXXXXXXX

(K) {20}

26: ILIXK KXXXXXXXXXXXXX

(K) {20}

27: LXXK KXXXXXXXXXXXXX

(K) {20}

28: XXXK KXXXXXXXXXXXXX

(K) {20}

29: KKKK KXXXXXXXXXXXXX

AAO10995 ck: 2382 len: 122 1 Aao10995 Human polypeptide SEQ ID NO 24887.

(R,K) {20,20}

(K) {20}

22: PCGGS KXXXXXXXXXXXXX

23: CKGSK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KGSKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
25: GSKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
26: SKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KGGGG
(K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK GGGGG
(K) {20}

AA010997 ck: 5315 len: 58 ! Aa010997 Human polypeptide SEQ ID NO 24889.
(R,K) {20,20}
(K) {20}

29: IXIIQ KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

30: XIIOK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

31: IIOKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}

32: IOKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}

33: OKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKGGP
(K) {20}

34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK RGGPP
(R,K) {20}

35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK GGPP
(R,K) {20}

AA011002 ck: 8638 len: 82 ! Aa011002 Human polypeptide SEQ ID NO 24894.
(R,K) {20,20}
(R,K) {20}

5: PRRS KKKKKKKKKKKKKKKKKKKKKKKKKR GGALL
(R,K) {20}

38: XFRRS KKKKKKKKKKKKKKKKKKKKKKKKKR GGALL
(R,K) {20}

AA011033 ck: 3780 len: 49 ! Aa011033 Human polypeptide SEQ ID NO 24925.
(R,K) {20,20}
(R,K) {20}

10: OKFTI RKKKKKKKKKKKKKKKKKKKKKKKK KLSKK
(K) {20}

11: XFFIR KKKKKKKKKKKKKKKKKKKKKKKKK LSKKI
(K) {20}

AA011048 ck: 7868 len: 85 ! Aa011048 Human polypeptide SEQ ID NO 24940.
(R,K) {20,20}
(K) {20}

53: IISDP KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

54: ISDPK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

55: SDPKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

56: DPKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

57: PKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

58: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

59: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

60: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

61: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

63: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}

64: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KK
(K) {20}

65: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK K
(K) {20}

66: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

AA011078 ck: 4186 len: 100 ! Aa011078 Human polypeptide SEQ ID NO 24970.
(R,K) {20,20}
(K) {20}

39: PPKPD KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

40: PKPDK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

41: KPDKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

42: PDKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

43: DKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

48: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKGKG
(K) {20}

1

50: KKKK (K) {20} KKKKKKKKKKKKKKKKK KRGKG
51: KKKK (K) {20} KKKKKKKKKKKKKKKKK RGGGG
52: KKKK (R,K) {20} KKKKKKKKKKKKKKKR GGGGF
AA01124 ck: 5121 len: 61 ! Aa01124 Human polypeptide SEQ ID NO 25016.
(R,K) {20,20}
16: KYSFL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
17: YSFLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: SFLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: FLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

36: KKKK (K) {20} KKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: KKKK (K) {20} KKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KKKK (K) {20} KKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: KKKK (K) {20} KKKKKKKKKKKKKKKKK KKK
(K) {20}
40: KKKK (K) {20} KKKKKKKKKKKKKKKKK KK
(K) {20}
41: KKKK (K) {20} KKKKKKKKKKKKKKKKK K
(K) {20}
42: KKKK (K) {20} KKKKKKKKKKKKKKKKK

AA01139 ck: 8807 len: 68 ! Aa01139 Human polypeptide SEQ ID NO 25031.

(R,K) {20,20}

(K) {20}

22: VCRFP KKKKKKKKKKKKKKKKKKK KMEGG

23: CRFPK KKKKKKKKKKKKKKKKKKK MEGGG

(K) {20}

1

AA01165 ck: 7203 len: 62 ! Aa01165 Human polypeptide SEQ ID NO 25057.

(R,K) {20,20}

(K) {20}

18: DASMV KKKKKKKKKKKKKKKKKKK KKKKK

19: ASMVK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

20: SMVKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

21: MVRKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

22: VKKKK KKKKKKKKKKKKKKKKKKK KKKKA

(K) {20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKAS

(K) {20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKAAS

(K) {20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KAASS

(K) {20}

26: KKKKK KKKKKKKKKKKKKKKKKKK ASSSQ

(K) {20}

AA01209 ck: 6562 len: 96 ! Aa01209 Human polypeptide SEQ ID NO 25101.

1

(R,K) {20,20}

(K) {20}

63: LLAYK KKKKKKKKKKKKKKKKKKK KKPXY

64: LLAYK KKKKKKKKKKKKKKKKKKK KPYLY

(K) {20}

(K) {20}

65: LAYKK KKKKKKKKKKKKKKKKKKKKKK PKYLS

AA011210 ck: 863 len: 70 1 Aa011210 Human polypeptide SEQ ID NO 25102.

(R,K){20,20}

(K){20}

31: DLCLC KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: LCLCK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

33: CLCLK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

34: LCKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

35: CKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKK PGGGG

AA011214 ck: 9584 len: 68 1 Aa011214 Human polypeptide SEQ ID NO 25106.

(R,K){20,20}

(K){20}

19: NRVCX KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: RVCXK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: VCXKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: CXKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: XKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKKKKK RGGPL

(R,K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKKKKK GGPLK

AA011241 ck: 1100 len: 36 1 Aa011241 Human polypeptide SEQ ID NO 25133.

(R,K){20,20}

(K){20}

8: XOTLT KKKKKKKKKKKKKKKKKKKKKK KAYTS

(K){20}

9: OTLTK KKKKKKKKKKKKKKKKKKKKKK AYTSK

AA011248 ck: 5013 len: 88 1 Aa011248 Human polypeptide SEQ ID NO 25140.

(R,K){20,20}

(K){20}

51: TCLYX KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: CLYXK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

53: LYXKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

54: YXKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

55: XKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

56: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

57: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKGGA

(K){20}

58: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKGAF

(K){20}

59: KKKKK KKKKKKKKKKKKKKKKKKKKKK GGAFK

AA011293 ck: 6490 len: 73 1 Aa011293 Human polypeptide SEQ ID NO 25185.

(R,K){20,20}

(K){20}

46: KXDYF KKKKKKKKKKKKKKKKKKKKKK KKS GG

(K){20}

47: XDYFK KKKKKKKKKKKKKKKKKKKKKK KKS GG

(K){20}

48: DYFKK KKKKKKKKKKKKKKKKKKKKKK KSGGG

(K){20}

49: YFKKK KKKKKKKKKKKKKKKKKKKKKK SGGGA

AA011324 ck: 6024 len: 93 1 Aa011324 Human polypeptide SEQ ID NO 25216.

(R,K){20,20}

(R,K){20}

37: ISLSA KKKKKKKKKKKKKKKKKKKKKK KKG GG

(K){20}

38: SLSAR KKKKKKKKKKKKKKKKKKKKKK KKG GG

(K){20}

39: LSARK KKKKKKKKKKKKKKKKKKKKKK KGG GG

40: SARRK (K) {20} KKKKKKKKKKKKKKKKKKKKK GGGGV
AA011327 ck: 3052 len: 102 ! Aa011327 Human polypeptide SEQ ID NO 25219.
(R, K) {20, 20}
(K) {20}
20: LSKEL KKKKKKKKKKKKKKKKKKKKK KKKLG
(K) {20}
21: SKELK KKKKKKKKKKKKKKKKKKKKK KKLGD
(K) {20}
22: KELKK KKKKKKKKKKKKKKKKKKKKK KLGDE
(K) {20}
23: ELKKK KKKKKKKKKKKKKKKKKKKKK LGDEE
AA011342 ck: 7189 len: 106 ! Aa011342 Human polypeptide SEQ ID NO 25234.
(R, K) {20, 20}
(K) {20}
61: FFKTX KKKKKKKKKKKKKKKKKKKKK GGPLK
AA011346 ck: 7311 len: 85 ! Aa011346 Human polypeptide SEQ ID NO 25238.
(R, K) {20, 20}
(K) {20}
62: ANHME KKKKKKKKKKKKKKKKKKKKK SGGG
AA011352 ck: 1342 len: 51 ! Aa011352 Human polypeptide SEQ ID NO 25244.
(R, K) {20, 20}
(K) {20}
25: SIFKX KKKKKKKKKKKKKKKKKKKKK KNPPS
(K) {20}
26: IFPKX KKKKKKKKKKKKKKKKKKKKK NPSSL
AA011705 ck: 7660 len: 101 ! Aa011705 Human polypeptide SEQ ID NO 25597.
(R, K) {20, 20}
(K) {20}
17: WAPLQ KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
18: APLQK KKKKKKKKKKKKKKKKKKKKK KKGGA
(K) {20}
19: PLQKK KKKKKKKKKKKKKKKKKKKKK KGGAP
(K) {20}
20: LQKKK KKKKKKKKKKKKKKKKKKKKK GGAPL
AA011799 ck: 8286 len: 39 ! Aa011799 Human polypeptide SEQ ID NO 25691.
(R, K) {20, 20}
(R, K) {20}
7: NPVVS KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
8: PPSVR KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
9: PVSRR KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

10: VSRKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: SRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
12: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGCG
(K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGGL
(K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK GGGL
AA011820 ck: 8725 len: 42 ! Aa011820 Human polypeptide SEQ ID NO 25712.
(R, K) {20, 20}
(K) {20}
17: NFKAL KKKKKKKKKKKKKKKKKKKKK SSSLR
AA011828 ck: 4807 len: 41 ! Aa011828 Human polypeptide SEQ ID NO 25720.
(R, K) {20, 20}
(K) {20}
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKRG
(K) {20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKRGG
(K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KRGGG
(K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK RGGGF
(R, K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK GGGEK
AA011831 ck: 1761 len: 29 ! Aa011831 Human polypeptide SEQ ID NO 25723.
(R, K) {20, 20}
(K) {20}
1: KKKKKKKKKKKKKKKKKKKKK KKKKA
(K) {20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKAG
(K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKAGG
(K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KAAGG

5: KKKK (K) {20}
AGGA

AA01843 ck: 86 len: 36 ! Aa01843 Human polypeptide SEQ ID NO 25735.

(R, K) {20, 20}

1: KKKK (K) {20}

2: K KKKK (K) {20}

3: KK KKKK (K) {20}

4: KKK KKKK (K) {20}

5: KKKK KKKK (K) {20}

6: KKKK KKKK (K) {20}

7: KKKK KKKK (K) {20}

8: KKKK KKKK (K) {20}

9: KKKK KKKK (K) {20}

10: KKKK KKKK (K) {20}

11: KKKK KKKK (K) {20}

12: KKKK KKKK (K) {20}

13: KKKK KKKK (K) {20}

14: KKKK KKKK (K) {20}

15: KKKK KKKK (K) {20}

AA01844 ck: 3574 len: 49 ! Aa01844 Human polypeptide SEQ ID NO 25736.

(R, K) {20, 20}

18: ILYME KKKK (K) {20}

19: LYMEK KKKK (K) {20}

20: YMEK KKKK (K) {20}

21: MEKKK KKKK (K) {20}

22: EKKKK KKKK (K) {20}

23: KKKK KKKK (K) {20}

24: KKKK KKKK (K) {20}

25: KKKK KKKK (K) {20}

26: KKKK KKKK (K) {20}

27: KKKK KKKK (K) {20}

AA01845 ck: 8375 len: 45 ! Aa01845 Human polypeptide SEQ ID NO 25737.

(R, K) {20, 20}

15: SPSTL KKKK (K) {20}

16: PSTLK KKKK (K) {20}

17: STLKK KKKK (K) {20}

18: TLKKK KKKK (K) {20}

19: LKKKK KKKK (K) {20}

20: KKKKK KKKK (K) {20}

21: KKKKK KKKK (K) {20}

22: KKKKK KKKK (K) {20}

23: KKKKK KKKK (K) {20}

24: KKKKK KKKK (K) {20}

AA01849 ck: 284 len: 36 ! Aa01849 Human polypeptide SEQ ID NO 25741.

(R, K) {20, 20}

13: KCLCE KKKK (K) {20}

14: CLCEK KKKK (R, K) {20}

15: LCEKK KKKK (R, K) {20}

16: CEKKK KKKK (R, K) {20}

17: EKKKK KKKK (R, K) {20}

AA01886 ck: 5282 len: 97 ! Aa01886 Human polypeptide SEQ ID NO 25778.

(R, K) {20, 20}

32: SNSSL KKKK (K) {20}

(K) {20}
33: NSSLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: SSLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: SLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AA011888 ck: 6488 len: 31 ! Aa011888 Human polypeptide SEQ ID NO 25780.
(R,K) {20,20}
(K) {20}
2: D KKKKKKKKKKKKKKKKKKKKK KAGGG
(K) {20}
3: DK KKKKKKKKKKKKKKKKKKKK AGGG

AA011903 ck: 8876 len: 48 ! Aa011903 Human polypeptide SEQ ID NO 25795.
(R,K) {20,20}
(K) {20}
17: VIKYV KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: IKYVV KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: KYVKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: YVKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

AA011999 ck: 3954 len: 47 ! Aa011999 Human polypeptide SEQ ID NO 25891.
(R,K) {20,20}
(K) {20}
17: IDEGL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: DEGLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: EGLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: GLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AA012098 ck: 8434 len: 39 ! Aa012098 Human polypeptide SEQ ID NO 25990.
(R,K) {20,20}
(K) {20}
13: TCKFG KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
14: CKFGK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
15: KFGKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
16: FGKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
17: GKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

19: KKKK (K) {20}
K
20: KKKK (K) {20}
K

AA012105 ck: 6255 len: 80 ! Aa012105 Human polypeptide SEQ ID NO 25997.

(R,K) {20,20}

(K) {20}

53: EATSL KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

54: ATSLK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

55: TSLK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

56: SLKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

57: LKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

58: KKKK KKKKKKKKKKKKKKKKKKK ARG

AA012179 ck: 1295 len: 69 ! Aa012179 Human polypeptide SEQ ID NO 26071.

(R,K) {20,20}

(K) {20}

35: PRESV KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

36: KPSVK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

37: FSVK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

38: SVKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

39: VKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

40: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

41: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

42: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

43: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

44: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

45: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

46: KKKK KKKKKKKKKKKKKKKKKKK GGGA

AA012180 ck: 9840 len: 67 ! Aa012180 Human polypeptide SEQ ID NO 26072.

(R,K) {20,20}

(K) {20}

38: CSYLP KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

39: SYLPK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

40: YLPK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

41: LPK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

42: PKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

43: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

44: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

45: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

46: KKKK KKKKKKKKKKKKKKKKKKK KK

(K) {20}

47: KKKK KKKKKKKKKKKKKKKKKKK K

(K) {20}

48: KKKK KKKKKKKKKKKKKKKKKKK

AA012187 ck: 4700 len: 60 ! Aa012187 Human polypeptide SEQ ID NO 26079.

(R,K) {20,20}

(K) {20}

30: SCKL KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

31: CKLK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

32: KCLK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

33: KCLK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

34: LKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

35: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

36: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

37: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

38: KKKK KKKKKKKKKKKKKKKKKKK KKKK

(K) {20}

39: KKKK KKKKKKKKKKKKKKKKKKK KK

(K) {20}

40: KKKK KKKKKKKKKKKKKKKKKKK K

(K) {20}

41: KKKK KKKKKKKKKKKKKKKKKKK


```
1
AA012203 ck: 4083 len: 41 i Aa012203 Human polypeptide SEQ ID NO 26095.
(R,K) {20,20}
6: SWCCL KKKKKKKKKKKKKKKKKKKKK KKPBG
(K) {20}
7: WCCLK KKKKKKKKKKKKKKKKKKKKK KKPBG
(K) {20}
8: CCLKK KKKKKKKKKKKKKKKKKKKKK KKPBG
(K) {20}
9: CLKKK KKKKKKKKKKKKKKKKKKKKK KPGGA
(K) {20}
AA012215 ck: 6903 len: 73 i Aa012215 Human polypeptide SEQ ID NO 26107.
(R,K) {20,20}
50: PPFLP KKKKKKKKKKKKKKKKKKKKK KTCGG
(K) {20}
51: PFLPK KKKKKKKKKKKKKKKKKKKKK TGG
(K) {20}
AA012243 ck: 8474 len: 39 i Aa012243 Human polypeptide SEQ ID NO 26135.
(R,K) {20,20}
11: MISPI KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
12: ISPIK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
13: SPIKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
14: FIKKK KKKKKKKKKKKKKKKKKKKKK KKGGP
(K) {20}
15: IKKKK KKKKKKKKKKKKKKKKKKKKK KGGPL
(K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK GGPL
(K) {20}
AA012250 ck: 1538 len: 69 i Aa012250 Human polypeptide SEQ ID NO 26142.
(R,K) {20,20}
31: SPSNL KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
32: PSNLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
33: SNLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: NLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
```

```
1
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG
(K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG
(K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK RGGGA
(K) {20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKK GGGAFF
(R,K) {20}
AA012274 ck: 2739 len: 37 i Aa012274 Human polypeptide SEQ ID NO 26166.
(R,K) {20,20}
14: KGLIN KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
15: GLINK KKKKKKKKKKKKKKKKKKKKK KKG
(K) {20}
16: LINKK KKKKKKKKKKKKKKKKKKKKK KG
(K) {20}
17: INKKK KKKKKKKKKKKKKKKKKKKKK G
(K) {20}
AA012280 ck: 7481 len: 66 i Aa012280 Human polypeptide SEQ ID NO 26172.
(R,K) {20,20}
29: HIPPL KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
30: IPPLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
31: PPLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
32: PLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
33: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKS
(K) {20}
```

1
39: KKKKK (K) {20}
40: KKKKK (K) {20}
41: KKKKK (K) {20}
42: KKKKK (K) {20}
AA012447 ck: 5017 len: 47 i Aa012447 Human polypeptide SEQ ID NO 26339.
(R, K) {20, 20}
(K) {20}
14: HLVC E KKKKKKKKKKKKKKKKKKKKK KRGGA
(K) {20}
15: LVCE KKKKKKKKKKKKKKKKKKKKK RGGAL
(R, K) {20}
16: VCEK KKKKKKKKKKKKKKKKKKKKK GGALK
AA012476 ck: 837 len: 104 i Aa012476 Human polypeptide SEQ ID NO 26368.
(R, K) {20, 20}
(K) {20}
65: RFCH Q KKKKKKKKKKKKKKKKKKKKK KGFFP
(K) {20}
66: FCHQ KKKKKKKKKKKKKKKKKKKKK GPFPP
AA012548 ck: 6973 len: 74 i Aa012548 Human polypeptide SEQ ID NO 26440.
(R, K) {20, 20}
(K) {20}
38: AVLP L KKKKKKKKKKKKKKKKKKKKK KKGCG
(K) {20}
39: VLPL KKKKKKKKKKKKKKKKKKKKK KGGGV
(K) {20}
40: LPLK KKKKKKKKKKKKKKKKKKKKK GGGVF
AA012553 ck: 1903 len: 33 i Aa012553 Human polypeptide SEQ ID NO 26445.
(R, K) {20, 20}
(K) {20}
9: SCCFI KKKKKKKKKKKKKKKKKKKKK KRGAP
(K) {20}
10: CCFI KKKKKKKKKKKKKKKKKKKKK RGAP
(R, K) {20}
11: CFIK KKKKKKKKKKKKKKKKKKKKK GAP
AA013164 ck: 2798 len: 71 i Aa013164 Human polypeptide SEQ ID NO 27056.
(R, K) {20, 20}
(R, K) {20}
35: RPPLX RKKKKKKKKKKKKKKKKKKKK KEMFK
(K) {20}
36: PPLX RKKKKKKKKKKKKKKKKKKKK EMFK
AA013576 ck: 4846 len: 99 i Aa013576 Human polypeptide SEQ ID NO 27468.

1
53: TNNLI (R, K) {20, 20}
(K) {20}
54: NNLI (K) {20}
55: NLI (K) {20}
56: LI (K) {20}
57: I (K) {20}
58: KKKK (K) {20}
59: KKKK (K) {20}
60: KKKK (K) {20}
AA013785 ck: 6241 len: 100 i Aa013785 Human polypeptide SEQ ID NO 27677.
(R, K) {20, 20}
(K) {20}
41: LYAP KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
42: YAP KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
43: APPK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
44: PPKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
45: PKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
46: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
47: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
48: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
49: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
50: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
51: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
52: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
53: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
54: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

55: KKKKK KKKKKKKKKKKKKKKKKKK KXAGG
(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KXAGG
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKK AGGGG
AAU17983 ck: 7016 len: 315 ! Aau17983 Human immunoglobulin polypeptide S
(R, K) {20, 20}
273: QVFAP RKKKKKKKKKKKKKKKKKK KGGRS
(R, K) {20}
274: VFAPR KKKKKKKKKKKKKKKKKKK GGRSR
(K) {20}
AAU18049 ck: 7611 len: 315 ! Aau18049 Human immunoglobulin polypeptide S
(R, K) {20, 20}
273: QVFAP RKKKKKKKKKKKKKKKKKK KGGRS
(R, K) {20}
274: VFAPR KKKKKKKKKKKKKKKKKKK GGRSR
(K) {20}
1
AAM14119 ck: 5383 len: 86 ! Aam14119 Peptide #553 encoded by probe for
(R, K) {20, 20}
15: RRRRG RRRRRKKKKKKKKKKKKKK KKRK
(R, K) {20}
16: RRRGR RRRKKKKKKKKKKKKKKKK KRRR
(R, K) {20}
17: RRGRR RRRKKKKKKKKKKKKKKKK KRRR
(R, K) {20}
18: RGRRR RKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
19: GRRRR RKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
20: RRRRR KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
21: RRRRK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
22: RRRKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
23: RRRKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
24: RKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
26: KKKKK RKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}

29: KKRKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
30: KKRKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
31: KKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
32: KKKKK RKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
33: KKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
34: KKKKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
35: KKRKK KKKKKKKKKKKKKKKKKKK RRRR
(R, K) {20}
36: KKKKK KRRRRKKKKKKKKKKKKKK RRRR
(R, K) {20}
37: RKKKK RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
38: KKKKK RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
39: KKKKK RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
40: KKKRR RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
41: KRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
42: RRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
43: RRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
44: RRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
45: RRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R) {20}
AAM14961 ck: 1334 len: 86 ! Aam14961 Peptide #1395 encoded by probe for
(R, K) {20, 20}
57: EEEEG RRRKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
58: EEEGR RRRKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
59: EEEGR RRRKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
60: EGGRR KKKKKKKKKKKKKKKKKKK KKKK
(R, K) {20}
61: GRRRK KRRRRKKKKKKKKKKKKKK KKKK
(R, K) {20}
62: RRRKK KRRRRKKKKKKKKKKKKKK KKKK
(R, K) {20}

63: RRRKK (R,K){20} KKKKK
64: RKKKK (R,K){20} KKKKK
65: KKKKK (R,K){20} KKKKK
66: KKKKK (R,K){20} K
67: KKKKK (K){20} KKKKK

AAM15038 ck: 9082 len: 167 1 Aam15038 Peptide #1472 encoded by probe for

33: EGRG (R,K){20,20} RRRRR
34: EGRG (R,K){20} RRRRR
35: GGRG (R,K){20} RRRRR

AAM15527 ck: 2276 len: 89 1 Aam15527 Peptide #1961 encoded by probe for

23: EEEE (R,K){20,20} KKKKK
46: KKEE (K){20} KKKKK
47: KKEE (K){20} KKKKK
48: EEEK (R,K){20} KKKKK
49: EEEK (R,K){20} KKKKK

AAM15826 ck: 1939 len: 130 1 Aam15826 Peptide #2260 encoded by probe for

42: EGKE (R,K){20,20} RRRRR
43: GRER (R,K){20} RRRRR
44: RKER (R,K){20} RRRRR
45: KERK (R,K){20} RRRRR
46: ERER (R,K){20} RRRRR
47: RRRR (R,K){20} RRRRR
48: RRRR (R,K){20} RRRRR

49: RRRR (R,K){20} RRRRR
50: RRRR (R,K){20} RRRRR
51: RRRR (R,K){20} RRRRR
52: RRRR (R,K){20} RRRRR
53: RRRR (R,K){20} RRRRR
54: RRRR (R,K){20} RRRRR

AAM16123 ck: 3607 len: 88 1 Aam16123 Peptide #2557 encoded by probe for

39: ERKE (R,K){20,20} KKKKK
40: RKEK (K){20} KKKKK
41: KEKK (K){20} KKKKK
42: REKK (K){20} KKKKK
43: EKKK (K){20} KKKKK
44: KKKK (K){20} KKKKK
45: KKKK (K){20} KKKKK
46: KKKK (K){20} KKKKK
47: KKKK (K){20} KKKKK
48: KKKK (K){20} KKKKK
49: KKKK (K){20} KKKKK
50: KKKK (K){20} KKKKK
51: KKKK (K){20} KKKKK
52: KKKK (K){20} KKKKK
53: KKKK (K){20} KKKKK
54: KKKK (K){20} KKKKK
55: KKKK (K){20} KKKKK

56: KKKK (K) {20}
57: KKKK (K) {20}
58: KKKK (K) {20}
59: KKKK (K) {20}
60: KKKK (K) {20}
61: KKKK (K) {20}
62: KKKK (K) {20}
63: KKKK (K) {20}
64: KKKK (K) {20}
65: KKKK (K) {20}
66: KKKK (K) {20}
67: KKKK (K) {20}
68: KKKK (K) {20}
69: KKKK (K) {20}

AAM1691 ck: 3937 len: 85 ! Aam1691 Peptide #3125 encoded by probe for

1
(R, K) {20, 20}
1: KKKK (K) {20}
2: K KKKK (K) {20}
3: KK KKKK (K) {20}
4: KK KKKK (K) {20}
5: KKKK (K) {20}
6: KKKK (K) {20}
7: KKKK (K) {20}
8: KKKK (K) {20}
9: KKKK (K) {20}

10: KKKK (K) {20}
11: KKKK (K) {20}

AAM1738 ck: 2686 len: 71 ! Aam1738 Peptide #3822 encoded by probe for

1
(R, K) {20, 20}
20: KKKK (K) {20}
21: KKKK (K) {20}
22: KKKK (K) {20}
23: KKKK (K) {20}
24: OKKK (K) {20}
25: KKKK (R, K) {20}
26: KKKK (R, K) {20}
27: KKKK (R, K) {20}
28: KKKK (R, K) {20}
29: KKKK (R, K) {20}
30: KKKK (R, K) {20}
31: KKKK (R, K) {20}
32: KKKK (R, K) {20}
33: KKKK (R, K) {20}
34: KKKK (R, K) {20}
35: KKKK (R, K) {20}
36: KKKK (R, K) {20}
37: KKKK (R, K) {20}
38: KKKK (R, K) {20}
39: KKKK (R, K) {20}
40: KKKK (R, K) {20}


```

22: RRRKK KKKKKKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
23: RRRKK KKKKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
24: RKKKK KKKKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
25: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
26: KKKKK RKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
27: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
28: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
29: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
30: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
31: RKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
32: KKKKK RKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
33: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
34: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
35: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
36: KKKKK KKKKKKKKKKKRRRR RRRRR
      (R, K) {20}
37: RKKKK RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
38: KKKKK RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
39: KKKKK RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
40: KKKKK RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
41: KKKKK RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
42: RRRRR RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
43: RRRRR RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
44: RRRRR RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}
45: RRRRR RRRRRRRRRRRRRRRRR RRRRR
      (R) {20}

```

```

1
(R,X){20,20}
57: EEERG RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
58: EEGR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
59: EEGRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
60: EGRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
61: GRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
62: RRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
63: RRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
64: RRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
65: RRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
66: RRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
67: RRRRR RRRKKKKKKKKKKKKKKKKKK
(R,K){20}

1
AAM27479 ck: 9082 len: 167 ! Aam27479 Peptide #1516 encoded by probe for
(R,X){20,20}
33: EERG RRRRRRRRRRRRRRRRRRRR
(R,K){20}
34: EGRG RRRRRRRRRRRRRRRRRRRR
(R,K){20}
35: GRRR RRRRRRRRRRRRRRRRRRRR
(R,K){20}

1
AAM28016 ck: 2276 len: 89 ! Aam28016 Peptide #2053 encoded by probe for
(R,X){20,20}
23: EEEE RRRRRRRRRRRRRRRRRRRR
(R,K){20}
46: KEEE RRRRRRRRRRRRRRRRRRRR
(R,K){20}
47: KEEE RRRRRRRRRRRRRRRRRRRR
(R,K){20}
48: EEEE RRRRRRRRRRRRRRRRRRRR
(R,K){20}
49: EEEE RRRRRRRRRRRRRRRRRRRR
(R,K){20}

1
AAM28336 ck: 1939 len: 130 ! Aam28336 Peptide #2373 encoded by probe for
(R,X){20,20}
42: EGRG RRRRRRRRRRRRRRRRRRRR
(R,K){20}

```

```

43: GKER RRRRRRRRRRRRRRRRRR (R,K){20}
44: RKER RRRRRRRRRRRRRRRRRR (R,K){20}
45: KER RRRRRRRRRRRRRRRRRR (R,K){20}
46: ERER RRRRRRRRRRRRRRRRRR (R,K){20}
47: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
48: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
49: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
50: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
51: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
52: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
53: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}
54: RRRR RRRRRRRRRRRRRRRRRR (R,K){20}

AAM28616 ck: 3607 len: 88 ! Aam28616 Peptide #2653 encoded by probe for
(R,K){20,20}
39: ERRE RRRRRRRRRRRRRRRRRR (K){20}
40: RREK RRRRRRRRRRRRRRRRRR (K){20}
41: REEK RRRRRRRRRRRRRRRRRR (K){20}
42: REKK RRRRRRRRRRRRRRRRRR (K){20}
43: EKKK RRRRRRRRRRRRRRRRRR (K){20}
44: KKKK RRRRRRRRRRRRRRRRRR (K){20}
45: KKKK RRRRRRRRRRRRRRRRRR (K){20}
46: KKKK RRRRRRRRRRRRRRRRRR (K){20}
47: KKKK RRRRRRRRRRRRRRRRRR (K){20}
48: KKKK RRRRRRRRRRRRRRRRRR (K){20}
49: KKKK RRRRRRRRRRRRRRRRRR (K){20}

```

```

50: KKKK (K) {20} KKKK
51: KKKK (K) {20} KKKK
52: KKKK (K) {20} KKKK
53: KKKK (K) {20} KKKK
54: KKKK (K) {20} KKKK
55: KKKK (K) {20} KKKK
56: KKKK (K) {20} KKKK
57: KKKK (K) {20} KKKK
58: KKKK (K) {20} KKKK
59: KKKK (K) {20} KKKK
60: KKKK (K) {20} KKKK
61: KKKK (K) {20} KKKK
62: KKKK (K) {20} KKKK
63: KKKK (K) {20} KKKK
64: KKKK (K) {20} KKKK
65: KKKK (K) {20} KKKK
66: KKKK (K) {20} KKKK
67: KKKK (K) {20} KKKK
68: KKKK (K) {20} KKKK
69: KKKK (K) {20} KKKK

AAM29178 ck: 3937 len: 85 | Aam29178 Peptide #3215 encoded by probe for
1: (R,K) {20,20} KKKK
2: (K) {20} KKKK
3: (K) {20} KKKK

```



```

4:   KKK {20}
    KKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
5:   KKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
6:   KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
7:   KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
8:   KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
9:   KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKEEE
      (K) {20}
10:  KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KEEEE
       (K) {20}
11:  KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK EEEEX

```

1

[illegible]

```

58: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR (R) {20}
59: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR (R) {20}
60: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR (R) {20}
61: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR (R) {20}
62: RRRR RRRRRRRRRRRRRRRRRRRRRRRRR (R) {20}
    NNNNE

AAM31173   ck: 2324   len: 36   i Aam31173 Peptide #5210 encoded by probe for
            (R,K){20,20}
16: KERKT KKKKKRRRRRRKKRRKKRRR R      (R,K){20}
17: ERKTK KKKKKRRRRRRKKRRKKRRRR      (R,K){20}

```

1

AAM33953 ck: 8343 len: 66 ! Aam33953 Peptide #7990 encoded by probe for

```

(R,K){20,20}
(R,K){20}
6: ETERE KKKKKKKKKKKKKKKKKKK KKKKK

```

(R,K) {20}

7: TEREK KKKKKKKKKKKKKKKKKKKKKKKKK

```

      (R,K) {20}
8: EREKK KKKKKKKKKKKKKKKKKKKKK KKKKK

```

```

(R,K) {20}
9: REKK KKKKKKKKKKKKKKKKKKK NKKKK

```

AAM36456 ck: 2394 len: 57 ! Aam36456 Peptide #10493 encoded by probe fo

1

```

(R,K){20,20}
(R){20}
20:EEEE RRRRRRRRRRRRRRRRR RRRRR

```

```

(R) { 20 }
21: EEGG RRRRRRRRRRRRRRRRR RRRG

```

(R) { 20 }

22: EGR RRRRRRRRRRRRRRRRR

(R) { 20 }

23 : EGR R R R R R R R R R R R R R R

24: GRRRR RRRRRRRRRRRR RGRRR (R) { 20 }

25: RRRR RRRRRRRRRRRR GRRR (R) { 20 }

AAM37018 ck: 4228 len: 24 ; Aam37018 Peptide #11055 encoded by probe fo

2

```

1:      (R,K) {20,20}
        (R,K) {20}
RRRRRRKKKKRRRRRRRR RRRT

```

2: R RRRRRKKKKRRRRRRRRR KRT (R, K) { 20}

3: RR RRRRRRRRRRRRRRRR RT
(R,K){20}
4: RRR RRRRRRRRRRRRRRRR T
(R,K){20}

AAM37794 ck: 3301 len: 52 i Aam37794 Peptide #11831 encoded by probe fd

12: KKKKN KKKKKKKKKKKKKKKK KKKK
(R,K){20,20}
13: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
14: KKNKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
15: KNKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
16: NKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
17: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
18: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
19: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
20: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
21: KKKKK RKKKKKKKKKKKKKKK KKKK
(R,K){20}
22: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
23: KKKKK RKKKKKKKKKKKKKKK KKKK
(R,K){20}
24: KKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
25: KKKKK RKKKKKKKKKKKKKKK KKKK
(R,K){20}
26: RKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
27: KKKKK RKKKKKKKKKKKKKKK KKKK
(R,K){20}
28: RKKKK KKKKKKKKKKKKKKKK KKKK
(R,K){20}
29: KKKKK KKKKKKKKKKKKKKKK KKKK
(K){20}
30: RKKKK KKKKKKKKKKKKKKKK KKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKK AF
(K){20}

AAM38273 ck: 4895 len: 51 i Aam38273 Peptide #12310 encoded by probe fd

18: LFKPM RKKRRKKRRKKKKKKRR KKLTT
(R,K){20,20}
19: FKPMR KRRKKKKRRKKKKRRK KLTTT
(R,K){20}
20: KPMRK RRRKKRRKKKKKKRRK LTTT
(R,K){20}

AAU04283 ck: 8137 len: 45 i Aau04283 Trimeric fusogenic peptide #2 used

4: YKA KKKKKKKKKKKKKKKKKK KKKK
(R,K){20,20}
5: YKAK KKKKKKKKKKKKKKKKK KKKK
(K){20}
6: YKAKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
7: KAKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
8: AKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}

1

23: KKKKK KKKKKKKKKKKKKKKKKKK KWK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KWK

AAU04285 ck: 4361 len: 59 ! Aau04285 Nuclear ligand #2 used in nucleic

(R,K) {20,20}

(K) {20}

18: APYKA KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

19: PYKAK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

20: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

21: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

22: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K) {20}

38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

AAU04287 ck: 4925 len: 100 ! Aau04287 Poly-L-Lysine used in nucleic acid
(R,K) {20,20}
(K) {20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

20: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

21: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

22: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

23: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK

24: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 25: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 26: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 27: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 28: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 29: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 30: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 31: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 32: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 33: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 34: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 35: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 36: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 37: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 38: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 39: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 40: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 41: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 42: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 43: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 44: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 45: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 46: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 47: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 48: KKKK KKKKKKKKKKKKKKKKKKK KKKK

49: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 50: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 51: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 52: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 53: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 54: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 55: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 56: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 57: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 58: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 59: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 60: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 61: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 62: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 63: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 64: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 65: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 66: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 67: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 68: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 69: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 70: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 71: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 72: KKKK KKKKKKKKKKKKKKKKKKK KKKK

73: KKKKK (K) {20} KKKKK
 74: KKKKK (K) {20} KKKKK
 75: KKKKK (K) {20} KKKKK
 76: KKKKK (K) {20} KKKKK
 77: KKKKK (K) {20} KKKKK
 78: KKKKK (K) {20} KKKKK
 79: KKKKK (K) {20} KKKKK
 80: KKKKK (K) {20} KKKKK
 81: KKKKK (K) {20} KKKKK

AAm01857 ck: 5383 len: 86 i Aam01857 Peptide #539 encoded by probe for

1

15: RRRRG (R,K) {20,20} RRRRG
 16: RRRGR (R,K) {20} RRRGR
 17: RRGRR (R,K) {20} RRGRR
 18: RGRRR (R,K) {20} RGRRR
 19: GRRRR (R,K) {20} GRRRR
 20: RRRRR (R,K) {20} RRRRR
 21: RRRRK (R,K) {20} RRRRK
 22: RRRKK (R,K) {20} RRRKK
 23: RRRKK (R,K) {20} RRRKK
 24: RRRKK (R,K) {20} RRRKK
 25: RRRKK (R,K) {20} RRRKK
 26: RRRKK (R,K) {20} RRRKK
 27: RRRKK (R,K) {20} RRRKK
 28: RRRKK (R,K) {20} RRRKK

1

29: KKKKK (R,K) {20} KKKKK
 30: KKKKK (R,K) {20} KKKKK
 31: KKKKK (R,K) {20} KKKKK
 32: KKKKK (R,K) {20} KKKKK
 33: KKKKK (R,K) {20} KKKKK
 34: KKKKK (R,K) {20} KKKKK
 35: KKKKK (R,K) {20} KKKKK
 36: KKKKK (R,K) {20} KKKKK
 37: KKKKK (R) {20} KKKKK
 38: KKKKK (R) {20} KKKKK
 39: KKKKK (R) {20} KKKKK
 40: KKKKK (R) {20} KKKKK
 41: KKKKK (R) {20} KKKKK
 42: KKKKK (R) {20} KKKKK
 43: KKKKK (R) {20} KKKKK
 44: KKKKK (R) {20} KKKKK
 45: KKKKK (R) {20} KKKKK

AAm02687 ck: 1334 len: 86 i Aam02687 Peptide #1369 encoded by probe for

57: EEEEG (R,K) {20,20} EEEEG
 58: EEEGR (R,K) {20} EEEGR
 59: EEEGR (R,K) {20} EEEGR
 60: EEEGR (R,K) {20} EEEGR
 61: EEEGR (R,K) {20} EEEGR
 62: EEEGR (R,K) {20} EEEGR

(R,K){20}
63: RRRKK KRRKKKKKKKKKKKKKKKKKK
(R,K){20}
64: RKKKK RRRKKKKKKKKKKKKKKKKKK
(R,K){20}
65: KKKKK KRRKKKKKKKKKKKKKKKKKK
(R,K){20}
66: KKKKK RRRKKKKKKKKKKKKKKKKKK
(K){20}
67: KKKRR KKKKKKKKKKKKKKKKKKK

1
AAM02768 ck: 9082 len: 167 1 Aam02768 Peptide #1450 encoded by probe for
(R,K){20,20}
33: EGGGG RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
34: EGGRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
35: GGGRR RRRRRRRRRRRRRRRRRRRRR

1
AAM03278 ck: 2276 len: 89 1 Aam03278 Peptide #1960 encoded by probe for
(R,K){20,20}
23: EEEEE KKKKKRRRRRRRRRRRRRRRR
(R,K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKKK
(K){20}
47: KEEKK KKKKKKKKKKKKKKKKKKKKK
(R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKKK
(R,K){20}
49: EEEKK KKKKKKKKKKKKKKKKKKKKK

1
AAM03564 ck: 1939 len: 130 1 Aam03564 Peptide #2246 encoded by probe for
(R,K){20,20}
42: EGGKE RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
43: GKKER RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
44: RKKER RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
45: KKKRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
46: EKKRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
47: RKKRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}

49: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRRR
(R,K){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRRR

1
AAM03848 ck: 3607 len: 88 1 Aam03848 Peptide #2530 encoded by probe for
(R,K){20,20}
39: EKKKE KKKKKKKKKKKKKKKKKKKKK
(K){20}
40: RKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
42: RKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K){20}

56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K) {20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK

1
AAM04408 ck: 3937 len: 85 ! Aam04408 Peptide #3090 encoded by probe for

(R, K) {20, 20}
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKK KEEEE
(K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK EEEEE

1
AAM06100 ck: 1560 len: 88 ! Aam06100 Peptide #4782 encoded by probe for

(R, K) {20, 20}
43: REERG RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
44: ERGRG RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
61: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
62: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

1
AAG73687 ck: 3063 len: 29 ! Aag73687 Human colon cancer antigen protein
(R, K) {20, 20}

(K) {20}
8: MMTTF KKKKKKKKKKKKKKKKKKKKK KX
(K) {20}
9: MMTFK KKKKKKKKKKKKKKKKKKKKK X

1
AAG73729 ck: 783 len: 83 ! Aag73729 Human colon cancer antigen protein

(R,K) {20,20}
(K) {20}
49: LGPCE KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
50: GPCEK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
51: PCEKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
52: CEKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
53: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

1
AAG73810 ck: 3374 len: 88 ! Aag73810 Human colon cancer antigen protein

(R,K) {20,20}
(K) {20}
44: FGQTX KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
45: GQTXK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
46: QTXKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
47: TXKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
48: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

1
AAG73895 ck: 1887 len: 43 ! Aag73895 Human colon cancer antigen protein

(R,K) {20,20}
(R,K) {20}

18: VRPRV RKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

19: RPRVR KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

20: PRVRK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

21: RVRKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

22: VRKKK KKKKKKKKKKKKKKKKKKKKK KKKK

1
AAG74218 ck: 8659 len: 104 ! Aag74218 Human colon cancer antigen protein

(R,K) {20,20}
(K) {20}

75: PLGGQ KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

76: LGGQK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

77: GGGQK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

78: GGGQK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

79: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

80: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

1

AAG74527 ck: 2664 len: 40 i Aag74527 Human colon cancer antigen protein
(R,K){20,20}
(K){20}
9: CULLY KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
10: LLLYK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
11: LLYK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
12: LYKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
13: YKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK

1

AAG74650 ck: 1596 len: 69 i Aag74650 Human colon cancer antigen protein
(R,K){20,20}
(K){20}
36: LQCRQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
37: QCRQK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
38: CROK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
39: RQKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
40: QKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
41: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
42: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
45: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
46: KKKK KKKKKKKKKKKKKKKKKKK KKKK

1

AAG74793 ck: 8497 len: 152 i Aag74793 Human colon cancer antigen protein
(R,K){20,20}
(K){20}
122: SHTQ KKKKKKKKKKKKKKKKKKK KKKK

1

123: SHTQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
AAG74907 ck: 1215 len: 98 i Aag74907 Human colon cancer antigen protein
(R,K){20,20}
(K){20}
57: NLRK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
58: LRKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
59: RKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
60: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
61: EKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
62: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
63: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
64: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
65: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
66: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
67: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
68: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
69: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
70: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
71: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
72: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
73: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
74: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
75: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
76: KKKK KKKKKKKKKKKKKKKKKKK KKKK

1

AAG75215 ck: 3913 len: 155 i Aag75215 Human colon cancer antigen protein
(R,K){20,20}
(K){20}

135: RSSAP KKKKKKKKKKKKKKKKKKK K
(K) {20}
136: SSAPK KKKKKKKKKKKKKKKKKKK

1
AAG75886 ck: 4235 len: 71 1 Aag75886 Human colon cancer antigen protein
(R,K) {20,20}
(K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1
AAE01796 ck: 4416 len: 72 1 Aae01796 Human gene 27 encoded secreted pro
(R,K) {20,20}
(K) {20}
47: LPTFL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: PTFLL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
49: TPLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
50: FLKKK KKKKKKKKKKKKKKKKKKK KKK

(K) {20}
51: LKKKK KKKKKKKKKKKKKKKKKKK KK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKK K
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKK

1
AAE01848 ck: 5584 len: 73 1 Aae01848 Human gene 27 encoded secreted pro
(R,K) {20,20}
(K) {20}
47: LPTFL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: PTFLL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
49: TPLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
50: FLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
51: LKKKK KKKKKKKKKKKKKKKKKKK KKK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKK K
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKK

(K) {20}
51: LKKKK KKKKKKKKKKKKKKKKKKK KKK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKK K

1
AAB90574 ck: 1431 len: 530 1 Aab90574 Human secreted protein, SEQ ID NO:
(R,K) {20,20}
(K) {20}
511: LHAPP KKKKKKKKKKKKKKKKKKK

1
AAB45846 ck: 8137 len: 45 1 Aab45846 Nucleic acid transporter system pei
(R,K) {20,20}
(K) {20}
4: YKA KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
5: YKAK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
6: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
8: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1
AAB45848 ck: 4361 len: 59 1 Aab45848 Nucleic acid transporter system pei
(R,K) {20,20}
(K) {20}
18: APYKA KKKKKKKKKKKKKKKKKKK KKKKK

[illegible]

4:	KKK	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
5:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
6:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
7:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
8:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
9:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
10:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
11:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
12:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
13:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
14:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
15:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
16:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
17:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
18:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
19:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
20:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
21:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
22:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
23:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
24:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
25:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
26:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	
27:	XXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXX
		(X) {20}	

77: KKKK (K) {20}
 78: KKKK (K) {20}
 79: KKKK (K) {20}
 80: KKKK (K) {20}
 81: KKKK (K) {20}

AAB50247 ck: 8085 len: 154 ! Aab50247 Human breast cancer associated B72

114: TOLRQ (R,K) {20,20}
 115: QLROK (K) {20}
 116: LRQK (K) {20}
 117: RQK (K) {20}
 118: QK (K) {20}
 119: KKKK (K) {20}
 120: KKKK (K) {20}
 121: KKKK (K) {20}
 122: KKKK (K) {20}
 123: KKKK (K) {20}
 124: KKKK (K) {20}
 125: KKKK (K) {20}
 126: KKKK (K) {20}
 127: KKKK (K) {20}
 128: KKKK (K) {20}

ABJ37738 ck: 8085 len: 154 ! Abj37738 Human tumour-related protein - SEQ
 (R,K) {20,20}
 114: TOLRQ (K) {20}
 115: QLROK (K) {20}

116: LRQK (K) {20}
 117: RQK (K) {20}
 118: QK (K) {20}
 119: KKKK (K) {20}
 120: KKKK (K) {20}
 121: KKKK (K) {20}
 122: KKKK (K) {20}
 123: KKKK (K) {20}
 124: KKKK (K) {20}
 125: KKKK (K) {20}
 126: KKKK (K) {20}
 127: KKKK (K) {20}
 128: KKKK (K) {20}

1 ABR00951 ck: 6732 len: 139 ! ABR00951 Human gene 5-encoded secreted prot
 (R,K) {20,20}
 91: HFRPG (K) {20}
 92: FRPG (K) {20}
 93: RPRG (K) {20}
 94: PRG (K) {20}
 95: GRG (K) {20}
 96: KKKK (K) {20}
 97: KKKK (K) {20}
 98: KKKK (K) {20}
 99: KKKK (K) {20}
 100: KKKK (K) {20}

[illegible]

```

44: VAKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKK
45: XKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKK
46: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK K
47: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK

AAE33919 ck: 185 len: 22 ! Aae33919 Secretion domain peptide #17. 5/20
1: (R,K) {20,20} (R) {20} RRRRRRRRRRRRRRRRRR GC

ABG73150 ck: 5536 len: 84 ! Abg73150 Single-chain antigen-binding (sFv)
1: (R,K) {20,20} (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
2: K (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
3: KK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
4: KKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
5: KKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

```

[illegible]

ABG73151 ck: 8024 len: 84 | Abg73151 Single-chain antigen-binding (sFv)

```

1      (R,K{20,20})
        (R){20}
RRRRRRRRRRRRRRRRRRRR RRRR
1 :
    (R){20}
R RRRRRRRRRRRRRRRRRRR RRRR
2 :
    (R){20}
```

[illegible]

- 24 : RRRR (R) {20} RRRRRRRRRRRRRRR
- 25 : RRRR (R) {20} RRRRRRRRRRRRRRR
- 26 : RRRR (R) {20} RRRRRRRRRRRRRRR
- 27 : RRRR (R) {20} RRRRRRRRRRRRRRR

28: RRRRR (R) {20}
 29: RRRRR (R) {20}
 30: RRRRR (R) {20}
 31: RRRRR (R) {20}
 32: RRRRR (R) {20}
 33: RRRRR (R) {20}
 34: RRRRR (R) {20}
 35: RRRRR (R) {20}
 36: RRRRR (R) {20}
 37: RRRRR (R) {20}

ABG73152 ck: 9643 len: 63 1 Abg73152 Single-chain antigen-binding (scFv)

1

1: RRRRR (R, K) {20, 20}
 2: RRRRR (R, K) {20}
 3: RRRRR (R, K) {20}
 4: RRRRR (R, K) {20}
 5: RRRRR (R, K) {20}
 6: RRRRR (R, K) {20}
 7: RRRRR (R, K) {20}
 8: RRRRR (R, K) {20}
 9: RRRRR (R, K) {20}
 10: RRRRR (R, K) {20}
 11: RRRRR (R, K) {20}
 12: RRRRR (R, K) {20}
 13: RRRRR (R, K) {20}

14: RRRRR (R, K) {20}
 15: RRRRR (R, K) {20}
 16: RRRRR (R, K) {20}
 17: RRRRR (R, K) {20}
 18: RRRRR (R, K) {20}
 19: RRRRR (R, K) {20}
 20: RRRRR (R, K) {20}
 21: RRRRR (R, K) {20}
 22: RRRRR (R, K) {20}
 23: RRRRR (R, K) {20}
 24: RRRRR (R, K) {20}
 25: RRRRR (R, K) {20}
 26: RRRRR (R, K) {20}
 27: RRRRR (R, K) {20}
 28: RRRRR (R, K) {20}
 29: RRRRR (R, K) {20}
 30: RRRRR (R, K) {20}
 31: RRRRR (R, K) {20}
 32: RRRRR (R, K) {20}
 33: RRRRR (R, K) {20}
 34: RRRRR (R, K) {20}
 35: RRRRR (R, K) {20}
 36: RRRRR (R, K) {20}
 37: RRRRR (R, K) {20}

[illegible]

```

24: RRRR RRRR{20} RRRR
25: RRRR RRRR{20} RRRR
26: RRRR RRRR{20} RRRR
27: RRRR RRRR{20} RRRR
28: RRRR RRRR{20} RRRR
29: RRRR RRRR{20} RRRR
30: RRRR RRRR{20} RRRR
31: RRRR RRRR{20} RRRR
32: RRRR RRRR{20} RRRR
33: RRRR RRRR{20} RRRR
34: RRRR RRRR{20} RRRR
35: RRRR RRRR{20} RRRR
36: RRRR RRRR{20} RRRR
37: RRRR RRRR{20} RRRR

ABG73870 ck: 5536 len: 84 ! Abg73870 Single-chain antigen-binding (SFV)
1:
(R,K){20,20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK
2: K{20} KKKKKKKKKKKKKKKKK
3: KK{20} KKKKKKKKKKKKKKKKK
4: KK{20} KKKKKKKKKKKKKKKKK
5: KK{20} KKKKKKKKKKKKKKKKK
6: KK{20} KKKKKKKKKKKKKKKKK
7: KK{20} KKKKKKKKKKKKKKKKK
8: KK{20} KKKKKKKKKKKKKKKKK
9: KK{20} KKKKKKKKKKKKKKKKK

```

10: KKKKK (K) {20}
11: KKKKK (K) {20}
12: KKKKK (K) {20}
13: KKKKK (K) {20}
14: KKKKK (K) {20}
15: KKKKK (K) {20}
16: KKKKK (K) {20}
17: KKKKK (K) {20}
18: KKKKK (K) {20}
19: KKKKK (K) {20}
20: KKKKK (K) {20}
21: KKKKK (K) {20}
22: KKKKK (K) {20}
23: KKKKK (K) {20}
24: KKKKK (K) {20}
25: KKKKK (K) {20}
26: KKKKK (K) {20}
27: KKKKK (K) {20}
28: KKKKK (K) {20}
29: KKKKK (K) {20}
30: KKKKK (K) {20}
31: KKKKK (K) {20}
32: KKKKK (K) {20}
33: KKKKK (K) {20}

34: KKKKK
35: KKKKK (K) {20}
36: KKKKK (K) {20}
37: KKKKK (K) {20}
ABG73871 ck: 8024 len: 84 ! AbG73871 Single-chain antigen-binding (sfv)
1: (R, K) {20, 20}
2: R (R) {20}
3: RR (R) {20}
4: RRR (R) {20}
5: RRRR (R) {20}
6: RRRRR (R) {20}
7: RRRRR (R) {20}
8: RRRRR (R) {20}
9: RRRRR (R) {20}
10: RRRRR (R) {20}
11: RRRRR (R) {20}
12: RRRRR (R) {20}
13: RRRRR (R) {20}
14: RRRRR (R) {20}
15: RRRRR (R) {20}
16: RRRRR (R) {20}
17: RRRRR (R) {20}
18: RRRRR (R) {20}
19: RRRRR (R) {20}

20: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
21: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
22: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
23: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
24: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
26: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
27: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
28: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
29: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
30: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
31: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
32: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
33: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
34: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
35: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
36: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}
37: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R) {20}

ABG73872 ck: 9643 len: 83 ! Abg73872 single-chain antigen-binding (sfv)
(R, K) {20, 20}
1: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
2: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
3: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
4: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
5: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

6: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
7: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
8: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
9: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
10: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
11: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
12: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
13: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
14: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
15: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
16: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
17: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
18: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
19: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
20: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
21: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
22: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
23: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
24: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
26: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
27: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
28: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
29: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}
30: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R, K) {20}

295: RRGKK (K) {20} KKKKK
 296: KGRKK (K) {20} KKKKK
 297: GKRRK (K) {20} KKKKK
 298: KKKKK (K) {20} KKKKK
 299: KKKKK (K) {20} KKKKK
 300: KKKKK (K) {20} KKKKK
 301: KKKKK (K) {20} KKKKK
 302: KKKKK (K) {20} KKKKA
 303: KKKKK (K) {20} KKKAP
 304: KKKKK (K) {20} KKAAP
 305: KKKKK (K) {20} KAAPP
 306: KKKKK (K) {20} APPPL
 342: PYKGE (K) {20} KKKKK
 343: YNGEK (K) {20} KKKKK
 344: KGEKK (K) {20} KKKKK
 345: GEEKK (K) {20} KKKKK
 346: EKKKK (K) {20} KKKKK
 347: KKKKK (K) {20} KKKKK
 348: KKKKK (K) {20} KKKKK
 349: KKKKK (K) {20} KKKKK
 350: KKKKK (K) {20} KKKKD
 351: KKKKK (K) {20} KKKDT
 352: KKKKK (K) {20} KKKDTL
 353: KKKKK (K) {20} KKKDTL

1

354: KKKKK (K) {20} KKKKK DTLTH
 ABJ19654 ck: 6732 len: 139 1 Abj19654 Human secreted protein amino acid
 (P,K) {20,20}
 91: HFRPG (K) {20} KKKKK
 92: FRPGK (K) {20} KKKKK
 93: RPRGK (K) {20} KKKKK
 94: PRGKK (K) {20} KKKKK
 95: GKRRK (K) {20} KKKKK
 96: KKKKK (K) {20} KKKKK
 97: KKKKK (K) {20} KKKKK
 98: KKKKK (K) {20} KKKKK
 99: KKKKK (K) {20} KKKKK
 100: KKKKK (K) {20} KKKKK
 101: KKKKK (K) {20} KKKKK
 102: KKKKK (K) {20} KKKKK
 103: KKKKK (K) {20} KKKKK
 104: KKKKK (K) {20} KKKKK
 105: KKKKK (K) {20} KKKKK
 106: KKKKK (K) {20} KKKKK
 107: KKKKK (K) {20} KKKKK
 108: KKKKK (K) {20} KKKKK
 109: KKKKK (K) {20} KKKKK
 110: KKKKK (K) {20} KKKKK
 111: KKKKK (K) {20} KKKKK
 112: KKKKK (K) {20} KKKKK

113: KKKK (K) {20} KKKK
 114: KKKK (K) {20} KKKK
 115: KKKK (K) {20} KKKK
 116: KKKK (K) {20} KKKK
 117: KKKK (K) {20} KKKK
 118: KKKK (K) {20} KKKK
 119: KKKK (K) {20} K
 120: KKKK (K) {20} KKKK

ABP99475 ck: 6732 len: 139 1 Abp99475 Human secreted protein SEQ ID NO 4

1

91: HPPG (R, K) {20, 20} KKKK
 92: FRPG (K) {20} KKKK
 93: RPGK (K) {20} KKKK
 94: PGKK (K) {20} KKKK
 95: GKKK (K) {20} KKKK
 96: KKKK (K) {20} KKKK
 97: KKKK (K) {20} KKKK
 98: KKKK (K) {20} KKKK
 99: KKKK (K) {20} KKKK
 100: KKKK (K) {20} KKKK
 101: KKKK (K) {20} KKKK
 102: KKKK (K) {20} KKKK
 103: KKKK (K) {20} KKKK
 104: KKKK (K) {20} KKKK
 105: KKKK (K) {20} KKKK

106: KKKK (K) {20} KKKK
 107: KKKK (K) {20} KKKK
 108: KKKK (K) {20} KKKK
 109: KKKK (K) {20} KKKK
 110: KKKK (K) {20} KKKK
 111: KKKK (K) {20} KKKK
 112: KKKK (K) {20} KKKK
 113: KKKK (K) {20} KKKK
 114: KKKK (K) {20} KKKK
 115: KKKK (K) {20} KKKK
 116: KKKK (K) {20} KKKK
 117: KKKK (K) {20} KKKK
 118: KKKK (K) {20} KKKK
 119: KKKK (K) {20} K
 120: KKKK (K) {20} KKKK

1

ABP9639 ck: 9531 len: 66 1 Abp9639 Human secreted protein SEQ ID NO 5f
 41: MMTVX (R, K) {20, 20} KKKK
 42: WTVXK (K) {20} KKKK
 43: TVXK (K) {20} KKKK
 44: VXXK (K) {20} KKKK
 45: XXXK (K) {20} KKKK
 46: KKKK (K) {20} K
 47: KKKK (K) {20} KKKK

Total finds: 8,655
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 07:06.66

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
!      1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V)
GENESQ2000S: AAB13781 ck: 1303 len: 25 finds: 1 ! Aab13781 Soluble peptide and
GENESQ2000S: AAB13783 ck: 4553 len: 45 finds: 1 ! Aab13783 Soluble tandem pEA/
\\End of list

Databases searched:
  Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds:      2
Total length:    158,726,570
Total sequences: 1,107,863
CPU time:        08:18.53
  
```



```

!!AA_SEQUENCE 1.0
ID AAB13781 standard; peptide; 25 AA.
XX
AC AAB13781;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble peptide antigen pEA.
XX
KW pEA peptide; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW major histocompatibility complex class I; MHC class I; antigen; tumour;
KW prostate; breast; multiple myeloma.
XX
XX Unidentified.
OS
XX WO20035949-A1.
XX
XX 22-JUN-2000.
XX
XX 14-DEC-1999; 99WO-US29724.
XX
XX 14-DEC-1998; 98US-0112324.
XX
XX (DEND-) DENDREON CORP.
XX
XX Laus R, Hakim I, Vidovic D;
XX
XX WPI; 2000-442365/38.
XX
XX Antigen modified by the covalent addition of a peptide that
XX facilitates entry into antigen presenting cells, useful for producing
XX compositions for immunizing against tumors and pathogens -
XX
XX Claim 2; Page 26; 34pp; English.
XX
XX The present invention relates to compositions of modified soluble protein
XX antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX response i.e. a major histocompatibility complex (MHC) class I molecule
XX response. The protein antigen is modified by the covalent addition of a
XX peptide sequence which facilitate entry of the antigen into antigen
XX presenting cells (APCs). The present sequence is one such peptide
XX sequence which can be used to modify the soluble antigens. The present
XX sequence is peptide pEA. The modified antigen composition may be used for
XX immunising against, or treating a tumour e.g. prostate and breast
XX carcinoma or multiple myeloma, or pathogen in mammals.
XX
XX Sequence 25 AA;
XX
AAB13781 Length: 25 January 30, 2004 10:59 Type: P Check: 1303 ..
1 CEAATAAATAA AAAAATAAAE AAAAA
!!AA_SEQUENCE 1.0
ID AAB13783 standard; peptide; 45 AA.
XX
AC AAB13783;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble tandem pEA/ pK peptide conjugate.
XX
KW pK peptide; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW major histocompatibility complex class I; MHC class I; antigen; tumour;
KW prostate; breast; multiple myeloma; pEA peptide.
XX
XX Unidentified.
OS
XX WO20035949-A1.
XX
XX 22-JUN-2000.
XX
XX 14-DEC-1999; 99WO-US29724.
XX

```

```

PR 14-DEC-1998; 98US-0112324.
XX
XX (DEND-) DENDREON CORP.
XX
XX Laus R, Hakim I, Vidovic D;
XX
XX WPI; 2000-442365/38.
XX
XX Antigen modified by the covalent addition of a peptide that
XX facilitates entry into antigen presenting cells, useful for producing
XX compositions for immunizing against tumors and pathogens -
XX
XX Claim 2; Page 26; 34pp; English.
XX
XX The present invention relates to compositions of modified soluble protein
XX antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX response i.e. a major histocompatibility complex (MHC) class I molecule
XX response. The protein antigen is modified by the covalent addition of a
XX peptide sequence which facilitate entry of the antigen into antigen
XX presenting cells (APCs). The present sequence is one such peptide
XX sequence which can be used to modify the soluble antigens. The present
XX sequence is tandem pEA/ pK peptide conjugate. The modified antigen
XX composition may be used for immunising against, or treating a tumour e.g.
XX prostate and breast carcinoma or multiple myeloma, or pathogen in
XX mammals.
XX
XX Sequence 45 AA;
XX
AAB13783 Length: 45 January 30, 2004 10:59 Type: P Check: 4553 ..
1 CEAATAAATAA AAAAATAAAE AAAAAKKKKK KKKKKKKKKK KKKKK

```


! FINDPATTERNS on pir:* allowing 0 mismatches

! 1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V)

Databases searched:

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds:

0

Total length:

96,168,682

Total sequences:

283,308

CPU time:

02:26.82

! FINDPATTERNS on swp:* allowing 0 mismatches

! 1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V)

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
 SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds:

0

Total length:

305,079,309

Total sequences:

958,388

CPU time:

08:14.71

! FINDPATTERNS on geneseqp: * allowing 0 mismatches

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5}

AAB13781 ck: 1303 len: 25 ! Aab13781 Soluble peptide antigen p5A. 11/20

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5}

2: C EAAAAAAAAAAAAAAAAA EAAAA

8: AAAAA EAAAAAAAAAAAAAAAAA

AAB13783 ck: 4553 len: 45 ! Aab13783 Soluble tandem p5A/ PK peptide con

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5}

2: C EAAAAAAAAAAAAAAAAA EAAAA

8: AAAAA EAAAAAAAAAAAAAAAAA KKKKK

ABP02760 ck: 5947 len: 86 ! Abp02760 Human ORFX protein sequence SEQ ID

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5}

13: GGRVR DACGCCEVCGALCGAVCG LQEGP

Databases searched:
Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 5
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 09:16.70


```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
!
! 1 (E,D) (A,L,I,F,G,C,M,V){5,5} (E,D) (A,L,I,F,G,C,M,V) {
GENESEQP2000S:AAB13781 ck: 1303 len: 25 finds: 2 ! Aab13781 Soluble peptide ant
GENESEQP2000S:AAB13783 ck: 4553 len: 45 finds: 2 ! Aab13783 Soluble tandem pEA/
GENESEQP2002S:ABP02760 ck: 5947 len: 86 finds: 1 ! ABP02760 Human ORFX protein
\\End of list

Databases searched:
Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 5
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 11:32.04

```



```

!!AA_SEQUENCE 1.0
ID AAB13781 standard; peptide; 25 AA.
XX
AC AAB13781;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble peptide antigen pEA.
XX
KM pEA peptide; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;
XX major histocompatibility complex class 1; MHC class 1; antigen; tumour;
XX prostate; breast; multiple myeloma.
XX
OS Unidentified.
XX
PN WO200035949-A1.
XX
PD 22-JUN-2000.
XX
PF 14-DEC-1999; 99WO-US29724.
XX
PR 14-DEC-1998; 98US-0112324.
XX
PA (DEND-) DENDREON CORP.
XX
PI Laus R, Hakim I, Vidovic D;
XX
DR WPI; 2000-442365/38.
XX
PT Antigens modified by the covalent addition of a peptide that
XX facilitates entry into antigen presenting cells, useful for producing
XX compositions for immunizing against tumors and pathogens -
XX
PS Claim 2; Page 26; 34pp; English.
XX
CC The present invention relates to compositions of modified soluble protein
XX antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX response i.e. a major histocompatibility complex (MHC) class I molecule
XX response. The protein antigen is modified by the covalent addition of a
XX peptide sequence which facilitate entry of the antigen into antigen
XX presenting cells (APCs). The present sequence is one such peptide
XX sequence which can be used to modify the soluble antigens. The present
XX sequence is peptide pEA. The modified antigen composition may be used for
XX immunizing against, or treating a tumour e.g. prostate and breast
XX carcinoma or multiple myeloma, or pathogen in mammals.
XX
SQ Sequence 25 AA;
AAB13781 Length: 25 January 30, 2004 11:00 Type: P Check: 1303 ..
1 CEAAMAAEAA AAAEAAAAAE AAAA
!!AA_SEQUENCE 1.0
ID AAB13783 standard; peptide; 45 AA.
XX
AC AAB13783;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble tandem pEA/ pK peptide conjugate.
XX
KM pK peptide; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;
XX major histocompatibility complex class 1; MHC class 1; antigen; tumour;
XX prostate; breast; multiple myeloma; pEA peptide.
XX
OS Unidentified.
XX
PN WO200035949-A1.
XX
PD 22-JUN-2000.
XX
PF 14-DEC-1999; 99WO-US29724.
XX

```

```

PR 14-DEC-1998; 98US-0112324.
XX
PA (DEND-) DENDREON CORP.
XX
PI Laus R, Hakim I, Vidovic D;
XX
DR WPI; 2000-442365/38.
XX
PT Antigens modified by the covalent addition of a peptide that
XX facilitates entry into antigen presenting cells, useful for producing
XX compositions for immunizing against tumors and pathogens -
XX
PS Claim 2; Page 26; 34pp; English.
XX
CC The present invention relates to compositions of modified soluble protein
XX antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX response i.e. a major histocompatibility complex (MHC) class I molecule
XX response. The protein antigen is modified by the covalent addition of a
XX peptide sequence which facilitate entry of the antigen into antigen
XX presenting cells (APCs). The present sequence is one such peptide
XX sequence which can be used to modify the soluble antigens. The present
XX sequence is tandem pEA/ pK peptide conjugate. The modified antigen
XX composition may be used for immunising against, or treating a tumour e.g.
XX prostate and breast carcinoma or multiple myeloma, or pathogen in
XX mammals.
XX
SQ Sequence 45 AA;
AAB13783 Length: 45 January 30, 2004 11:00 Type: P Check: 4553 ..
1 CEAAMAAEAA AAAEAAAAAE AAAAAXKKKK KKKKKKKKK KKKKK
!!AA_SEQUENCE 1.0
ID AAB02760 standard; Protein; 86 AA.
XX
AC AAB02760;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:5502.
XX
KM Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
XX
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach MD;
XX
DR N-PSDB; AEN18512.
XX
PT Novel human polypeptides and polymucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 5502; 1037pp; English.

```

XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). AEN15162 to AEN27752 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating
CC a syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration or treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 86 AA;

ABP02760 Length: 86 January 30, 2004 11:00 Type: P Check: 5947 .

1 XRPPTDCEGR VRDACCCEV CGALEGAVCG LQEGPCGEA ANAVSAPPSG

51 VPASATVRRR AQAQGLCVCCAS SEPVCNDK TYTNLC

! FINDPATTERNS on swp:* allowing 0 mismatches

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {

HRAL_MOUSE ck: 7206 len: 480 ! Q9r118 mus musculus (mouse). serine protease

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C) {5}

60: GGRVR DACGCCVCGALGGAACG LQEGP

1 Q9KHD9 ck: 1697 len: 262 ! Q9khd9 streptomyces griseus subsp. griseus.

(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(E) (A,L,C) {5} (D) (A,L,G,V) {5} (E) (A,L,F,G) {5}

124: LGGGL ELALACDLVAVGALFG LPBLG

1 Q8RUS6 ck: 3612 len: 107 ! Q8rus6 oryza sativa (japonica cultivar-grou

(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(E) (A,I,V) {5} (E) (G,C,V) {5} (D) (A,L,G,M,V) {5}

20: LETEV EAIAVEGGCGVDLAVG RALGL

1 Q9QZK6 ck: 7594 len: 480 ! Q9qzk6 mus musculus (mouse). insulin-like g

(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C) {5}

60: GGRVR DACGCCVCGALGGAACG LQEGP

1 Q91WS3 ck: 7882 len: 480 ! Q91ws3 mus musculus (mouse). protease, seri

(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C) {5}

60: GGRVR DACGCCVCGALGGAACG LQEGP

1 Q9QZK5 ck: 8689 len: 480 ! Q9qzk5 rattus norvegicus (rat). insulin-lik

(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C,V) {5}

60: GGRVR DACGCCVCGALGGAACG LQEGP

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 6

Total length: 305,079,309

Total sequences: 958,388

CPU time: 14:58.50


```
!!SEQUENCE LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches
!
! 1 (E,D) (A,L,I,F,G,C,M,V){5,5} (E,D) (A,L,I,F,G,C,M,V) {
```

```
SW:HRAL_MOUSE      ck: 7206 len: 480 finds: 1 ! Q9r118 mus musculus (mouse) . S
SP_BA:Q9KHD9        ck: 1697 len: 262 finds: 1 ! Q9khd9 streptomyces griseus su
SP_PU:Q8RUS6        ck: 3612 len: 107 finds: 1 ! Q8rus6 oryza sativa (japonica)
SP_RO:Q9QZK6        ck: 7594 len: 480 finds: 1 ! Q9qzk6 mus musculus (mouse) . I
SP_RO:Q91WS3        ck: 7882 len: 480 finds: 1 ! Q91ws3 mus musculus (mouse) . P
SP_RO:Q9QZK5        ck: 8689 len: 480 finds: 1 ! Q9qzk5 rattus norvegicus (rat)

\\End of list
```

```
Databases searched:
SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPRTRMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003
```

```
Total finds: 6
Total length: 305,079,309
Total sequences: 958,388
CPU time: 16:49.45
```



```

!!AA_SEQUENCE 1.0
ID HRA1_MOUSE STANDARD; PRT; 480 AA.
AC O9R18;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease HTRA1 precursor (EC 3.4.21.-).
GN PRS11 OR HTRA1 OR HTRA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RA Oka C., Soma A., Kanda H., Kawachi M.;
RT "The role of murine serine protease HTRA in osteogenesis.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Protease that regulate the availability of IGFs by
CC cleaving IGF-binding proteins (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- SIMILARITY: BELONGS TO PROTEINASE FAMILY S2C.
CC -1- SIMILARITY: Contains 1 IGFBP domain.
CC -1- SIMILARITY: Contains 1 Kazal-like domain.
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF172994; AAD9422.1; -.
DR MEROPS; S01.277; -.
DR MGD; MGI:1929076; Prs11.
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR InterPro; IPR002350; Kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00050; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASE2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolyase; Serine protease; Growth factor binding; Signal.
FT SIGNAL 1 22
FT CHAIN 23 480
FT DOMAIN 37 94
FT DOMAIN 101 155
FT DOMAIN 204 364
FT DOMAIN 365 467
FT ACT_SITE 220 220
FT ACT_SITE 250 250
FT ACT_SITE 328 328
SQ SEQUENCE 480 AA; 51246 MW; 54BB9BA6C99A7BFA CRC64;

HRA1_MOUSE Length: 480 January 30, 2004 15:08 Type: P Check: 7206
1 MOSLRTTLLS LLLILLAPS LALPSGTGRS APAATVCPDH CDPTRCAPP
51 TDCSGRVND ACGCEVCGA LEGACGLQIE GPCGELQCV LPFGVPASAT
101 VRRRAQAGLC VCASPEVCG SDAKTNNLC QLRASRRE KLPPQPVIVL
151 ORGACGQGOE DPNLRHKYN FIADVVEKFA PDVVKHELVR KLPPSKREVP
201 VASGGFTVS EDGLIVTNAH VTNKRVKV ELKNGATYEA IIKVDEKAD

```

```

!!AA_SEQUENCE 1.0
ID Q9KHD9 PRELIMINARY; PRT; 262 AA.
AC Q9KHD9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Enoyl-CoA hydratase-like protein.
OS Streptomyces griseus subsp. griseus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=67263;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM40695;
RA MEDLINE=20316791; PubMed=10858335;
RX Smith W.C., Xiang L., Shen B.;
RT "Genetic localization and molecular characterization of the nons gene
RT required for macroretrolyde biosynthesis in Streptomyces griseus
RT DSM40695.";
RL Actinmicrob. Agents Chemother. 44:1809-1817(2000).
DR EMBL; AF263011; AAF81231.1; -.
DR HSSP; P14604; 2DUB.
DR InterPro; IPR001753; EnCoA_hydrase.
DR Pfam; PF00378; ECH; 1.
DR PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
SQ SEQUENCE 262 AA; 27068 MW; D218PA2BDEC64BF2 CRC64;

Q9KHD9 Length: 262 January 30, 2004 15:08 Type: P Check: 1697
1 MTPQETAPBEG EPEPLIVERH GRVATLTLLNR PHRRNANSTA MLARLDHALG
51 KLITQGGSEAP GALVLTGAGG TFSGGADTRE PDWEDLSRRA VRRAPFTVF
101 AMHEAPPPV VAAVEGYALG GGLIELALCD LVVAGEGALF GLPELGVAV
151 PGGAHVSLV RRAAGRYAAR MLITGEERYR ADELARLGA VERTVDDGAL
201 AEAQALASV AAGDPALLBA GVRLLRDSGH LDRTPALGVE NGYWMQASAA
251 ANRDPSPSSG RP

!!AA_SEQUENCE 1.0
ID Q9RUS6 PRELIMINARY; PRT; 107 AA.
AC Q9RUS6;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE P0678F11.3 protein (P0413C03.30 protein).
GN P0678F11.3 OR P0413C03.30.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0678F11.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
   clone: p0413C03."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003437; BAB86096.1; -
DR EMBL; AP003451; BAB86154.1; -
DR Gramene; Q8RUS6; -
SQ SEQUENCE 107 AA; 11046 MW; 89097D88201BD849 CRC64;

Q8RUS6 Length: 107 January 30, 2004 15:08 Type: P Check: 3612 ..

1 MESPRAVNL RMSLETEVE AALAVEGCG VLAAMGRAL GLDPATVALN
51 GYFVRGRGH VSAVTWRAL LDFPARGLP TGDAPAPVA VHGRAPPPP
101 PPPVSVL

11AA SEQUENCE 1.0
ID _Q9QZK6 PRELIMINARY; PRT; 480 AA.
AC _Q9QZK6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Insulin-like growth factor binding protein 5 protease.
GN PRS11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Houtvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
   Roby J.A., Mayo K.E., Adashi E.Y.;
RT "Mouse insulin-like growth factor binding protein 5-directed
   endopeptidase: structural assessment, evolutionary analysis, ovarian
   expression, hormonal regulation and cellular localization."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; AF179369; AAD52682.1; -
DR MGD; MGI:1929076; Prs11.
DR InterPro; IPR000867; Ins1_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00219; IGFBP; 1.
DR Pfam; PF00505; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 480 AA; 51213 MW; 92BDDA85CF5B1287 CRC64;

Q9QZK6 Length: 480 January 30, 2004 15:08 Type: P Check: 7594 ..

1 MSLRTTLLS LLLLLLAAPS LALPSGTGRS APAATVCPFH CDPTRCAPPP
51 TDCEGRVRD AGCCCEVCGA LEGAAGCLQE GPCGEGLOCV VPGVPASAT
101 VRRRAQGLC VCASSEPVCG SDAKTYTNC QLRASRSE KLRPPVIVL
151 QRGACGQGE DPNLSRHKN FIAVVEKIA PAVVHIELYR KLPSKREVP
201 VASGSGFIVS EDGLIVTNAH VTNKNRKYV ELKNGATVEA KIKDVEDKAD
251 IALIKIDHOG KLPVLLIGRS SELRGEFV AIGSPFSLQN TVTTGIVSTT

```

```

301 QRGKELGLR NSDMYIQT DAIINYNSGG PLVNLDEVI GINTLKVTAG
351 ISFALPSKI KKFLETSHR QAKKAVTKK KYIGRMSL TSSAKELKD
401 RHRDPDVLG GAVIIEVIPD TPAEAGLKE NDVIISINGQ SVTANDVSD
451 VIKKENTLIM VVRGNEDIY ITVPEIDP

11AA SEQUENCE 1.0
ID _Q91WS3 PRELIMINARY; PRT; 480 AA.
AC _Q91WS3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Protease, serine, 11 (1gf binding).
GN PRS11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strauberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; BC013516; AAH13516.1; -
DR MGD; MGI:1929076; Prs11.
DR InterPro; IPR000867; Ins1_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00219; IGFBP; 1.
DR Pfam; PF00595; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 480 AA; 51212 MW; 76BDD5B862EDC9DA CRC64;

Q91WS3 Length: 480 January 30, 2004 15:08 Type: P Check: 7882 ..

1 MSLRTTLLS LLLLLLAAPS LALPSGTGRS APAATVCPFH CDPTRCAPPP
51 TDCEGRVRD AGCCCEVCGA LEGAAGCLQE GPCGEGLOCV VPGVPASAT
101 VRRRAQGLC VCASSEPVCG SDAKTYTNC QLRASRSE KLRPPVIVL
151 QRGACGQGE DPNLSRHKN FIAVVEKIA PAVVHIELYR KLPSKREVP
201 VASGSGFIVS EDGLIVTNAH VTNKNRKYV ELKNGATVEA KIKDVEDKAD
251 IALIKIDHOG KLPVLLIGRS SELRGEFV AIGSPFSLQN TVTTGIVSTT
301 QRGKELGLR NSDMYIQT DAIINYNSGG PLVNLDEVI GINTLKVTAG
351 ISFALPSKI KKFLETSHR QAKKAVTKK KYIGRMSL TSSAKELKD
401 RHRDPDVLG GAVIIEVIPD TPAEAGLKE NDVIISINGQ SVTANDVSD
451 VIKKENTLIM VVRGNEDIY ITVPEIDP

```

```

11AA SEQUENCE 1.0
ID _Q9QZK5 PRELIMINARY; PRT; 480 AA.
AC _Q9QZK5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

```

DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Insulin-like growth factor binding protein 5 protease.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Ovary;
 RA Hourvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
 RA Roby J.A., Mayo K.E., Adashi E.Y.;
 RT "Mouse Insulin-like growth factor binding protein 5-directed
 RT endopeptidase: structural assessment, evolutionary analysis, ovarian
 RT expression, hormonal regulation and cellular localization";
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL: AF179370; AAD52683.1; -.
 DR MEROPS: S01.277; -.
 DR InterPro: IPR000867; Ins1_gro_fac_pr.
 DR InterPro: IPR002350; Kazal.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR01254; Ser_protease_Try.
 DR Pfam: PF00219; IGFBP_1.
 DR Pfam: PF00050; Kazal; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00121; IB; 1.
 DR SMART: SM00280; KAZAL; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 KM Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 480 AA; 5130 MW; 37A864C5A8FC035 CRC64;
 Q9QZK5 length: 480 January 30, 2004 15:08 Type: P Check: 8689 ..
 1 MQFLRTALLS LLLLLLAAPS LALPSGISRS APAATVCPFH CDPTRCAPP
 51 TDCSGGRVRD ACGCEVCGA LBGAVCGLGE GPCGEGLOCV VPGVVPASAT
 101 VRRRAQAGLC VCASSEPCVG SDAKTYTNLC QLRASRSE KLRQPPVIVL
 151 QRGACGQGOE DPNGLRHKYN FIADVETKIA PAVVHIELYR KLPSKREVP
 201 VASGSGFIVS EDGLIVTNAH VTNKRVKY ELKNGATTEA KIKVDEKAD
 251 IALIKIDHOG KLPVLLGRS SELRPGFVV AIGSPFSLQN TVTTGIIVSTT
 301 QRGKELGLR NSDMDIQTD AINYNSSG PLVNLGDEVI GINTLKVTAG
 351 ISFALPSDKI KKLFTSHDR QAKGTVTKK KYIGIRMSL TSSKAKELKD
 401 RHRDEPDVIS GAYLIEVIPD TPAAGGLKE NDVIISINQ SVVTANDVSD
 451 VIKKENTLNM VVRGNEDIV ITVPEEIDP

! FINDPATTERNS on pir:* allowing 0 mismatches

! 1 (E,D) (A,L,I,F,G,C,M,V){5,5} (E,D) (A,L,I,F,G,C,M,V) {

Databases searched:

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds:

0

Total length:

96,168,682

Total sequences:

283,308

CPU time:

04:36.11

THIS PAGE BLANK (USPTO)